AstraZeneca

Statistical Analysis Plan

Study Code	D4191C00001
Edition Number	4.0
Date	22 March 2017

A Phase III, Randomised, Double-blind, Placebo-controlled, Multi-centre, International Study of MEDI4736 as Sequential Therapy in Patients with Locally Advanced, Unresectable Non-Small Cell Lung Cancer (Stage III) Who Have Not Progressed Following Definitive, Platinum-based, Concurrent Chemoradiation Therapy (PACIFIC)

A Phase III, Randomised, Double-blind, Placebo-controlled, Multi-centre, International Study of MEDI4736 as Sequential Therapy in Patients with Locally Advanced, Unresectable Non-Small Cell Lung Cancer (Stage III) Who Have Not Progressed Following Definitive, Platinum-based, Concurrent Chemoradiation Therapy (PACIFIC)

Study Statistician

PPD

Lynne Poole

22 March 2017.

Date

A Phase III, Randomised, Double-blind, Placebo-controlled, Multi-centre, International Study of MEDI4736 as Sequential Therapy in Patients with Locally Advanced, Unresectable Non-Small Cell Lung Cancer (Stage III) Who Have Not Progressed Following Definitive, Platinum-based, Concurrent Chemoradiation Therapy (PACIFIC)

Global Product Statistician PPD

Yifan Huang

3/22/2017
Date

TABLE OF CONTENTS PAGE

TITLE PAGE 1
SIGNATURE OF STUDY STATISTICIAN 2
SIGNATURE OF GLOBAL PRODUCT STATISTICIAN 3
TABLE OF CONTENTS 4
LIST OF ABBREVIATIONS 8
AMENDMENT HISTORY 11
1. STUDY DETAILS 14
1.1 Study Objectives 14
1.1.1 Primary objective 14
1.1.2 Secondary objectives 14
1.1.3 Exploratory objectives 15
1.2 Study Design 16
1.3 Number of Subjects 21
2. ANALYSIS SETS 21
2.1 Definition of Analysis Sets 21
2.2 Protocol Deviations 23
3. PRIMARY, SECONDARY AND EXPLORATORY VARIABLES 25
3.1 Derivation of RECIST Visit Responses 25
3.1.1 Blinded Independent Central Review (BICR) Assessment Using RECIST 1.1 26
3.1.2 BICR assessment using RECIST 1.1 Modified for Confirmation of Progression 27
3.1.3 BICR assessment using irRECIST 1.1 28
3.1.4 Site investigator assessment using RECIST 1.1: target lesions 28
3.1.5 Site investigator assessment using RECIST 1.1: NTLs and new lesions 33
3.1.6 Site investigator assessment using RECIST 1.1: overall visit response 34
3.2 Outcome Variables 35
3.2.1 Co-primary efficacy outcome endpoints 35
3.2.1.1 Overall survival (OS) 35
3.2.1.2 Progression free survival (PFS) 36
3.2.2 Secondary efficacy outcome endpoints 38
3.2.2.1 Proportion of patients alive at 24 months 38
3.2.2.2 Objective response rate 38

3.2.2.3 Duration of response 40
3.2.2.4 Proportion of patients alive and progression free at 12 months 40
3.2.2.5 Proportion of patients alive and progression free at 18 months 40
3.2.2.6 Time from randomisation to second progression or death (PFS2) 41
3.2.2.7 Time to death or distant metastasis 41

3.3 Secondary Patient Reported Outcome Variables 42
3.3.1 EORTC-QLQ-C30 42
3.3.1.1 Time to symptom deterioration 43
3.3.1.2 Time to HR QoL/Function deterioration 44
3.3.1.3 Symptom Improvement Rate 45
3.3.1.4 HR QoL/Function Improvement Rate 45
3.3.2 EORTC-QLQ-LC-13 45
3.3.2.1 Time to symptom deterioration 46
3.3.2.2 Symptom Improvement Rate 47
3.3.3 PRO Compliance Rates 47

3.4 Safety 48
3.4.1 Adverse events 48
3.4.2 Laboratory data 49
3.4.3 ECGs 50
3.4.4 Vital signs 50
3.4.5 Time to first subsequent therapy from discontinuation of study treatment 51
3.4.6 Treatment exposure 51
3.4.7 Dose intensity 52

3.5 Pharmacokinetic and Immunogenicity Variables 53
3.5.1 PK non-compartmental analysis 53
3.5.2 Population PK and exposure-response/safety analysis 53
3.5.3 Immunogenicity analysis 53

3.6 Exploratory Variables 54
3.6.1 EQ-5D-5L 54
3.6.2 Health Resource Use 54

4. ANALYSIS METHODS 55

4.1 General Principles 55
4.1.1 Baseline 56
4.1.2 Multiplicity 56
4.1.3 Visit window for safety and PRO assessments 59

4.2 Analysis Methods 61
4.2.1 Co-primary efficacy endpoints 63
4.2.1.1 Overall survival 63
4.2.1.2 Progression free survival 67
4.2.2 Secondary efficacy endpoints 70
4.2.2.1 Proportion of patients alive at 24 months (OS24) 70
4.2.2.2 Objective response rate 71

4.2.2.3 Duration of response 72
4.2.2.4 Proportion of patients alive and progression free at 12 months (APF12) 72
4.2.2.5 Proportion of patients alive and progression free at 18 months (APF18) 72
The proportion of patients alive and progression free at 18 months (i.e. at study day 549) will be summarised (using the Kaplan-Meier curve) and presented by treatment arm. For each treatment arm, the APF18 based on Kaplan-Meier method will be presented, along with its 95% confidence interval. The computation of the confidence interval will be based on a log(-log(.)) transformation. 72
4.2.2.6 Time from randomisation to second progression (PFS2) 72
4.2.2.7 Time to death or distant metastasis 73
4.2.2.8 Change in tumour size 73
4.2.3 Secondary patient reported outcome endpoints 73
4.2.3.1 EORTC QLQ-C30 74
4.2.3.2 EORTC-QLQ-LC13 74
4.2.3.3 Mixed models repeated measures of change from baseline in PRO symptoms 75
4.2.4 Safety data 77
4.2.4.1 Adverse events 77
4.2.4.2 Laboratory assessments 80
4.2.4.3 ECGs 82
4.2.4.4 Vital signs 82
4.2.4.5 Physical examination 82
4.2.4.6 Other safety data 82
4.2.5 WHO performance status 82
4.2.6 PK analysis 82
4.2.7 Immunogenicity analysis 83
4.2.8 PK/PDx relationships 83
4.2.9 Exploratory endpoints 83
4.2.9.1 EQ-5D-5L 83
4.2.9.2 Health resource use 83
4.2.10 Demographic and baseline characteristics data 84
4.2.11 Treatment exposure and intensity 85

5. INTERIM ANALYSES 85

5.1 Analysis Methods 85
5.1.1 PFS interim analysis 86
5.1.2 OS interim analyses 86

5.2 Independent Data Monitoring Committee 87

6. CHANGES OF ANALYSIS FROM PROTOCOL 89

7. REFERENCES 90

8. APPENDIX – DETAILS FOR TTDM 94

LIST OF TABLES

Table 1 Summary of outcome variables and analysis populations....22
Table 2 TL Visit Responses....29
Table 3 NTL Visit Responses....33
Table 4 Overall Visit Responses....34
Table 5 Visit response for EORTC-QLQ-C30....43
Table 6 Visit response for L13....46
Table 7 Dose intensity scenarios....52
Table 8 Formal statistical analyses to be conducted and pre-planned sensitivity analyses....61
Table 9 Changes from the planned analysis....89

LIST OF FIGURES

Figure 1 Study flow chart....19
Figure 2 Multiple testing procedures for controlling the type 1 error rate....59

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
ADA	Anti-drug antibody
AE	Adverse event
AESI	Adverse events of special interest
ALT	Alanine aminotransferase
APF12	Proportion of patients alive and progression free at 12 months from randomisation
APF18	Proportion of patients alive and progression free at 18 months from randomisation
AST	Aspartate aminotransferase
AZDD	AZ drug dictionary
Baseline	Refers to the most recent assessment of any variable prior to dosing with study treatment/randomisation (as appropriate)
BICR	Blinded independent central review
BoR	Best overall RECIST response
CI	Confidence interval
CR	Complete response
CRF / eCRF	Case Report Form (electronic)
CRO	Contract Research Organisation
CSP	Clinical Study Protocol
CSR	Clinical Study Report
CT	Computed tomography
CTCAE	Common Terminology Criteria for Adverse Event
DAE	Discontinuation of investigational product due to adverse event
DBL	Data base lock
DBP	Diastolic blood pressure
DCO	Data cut-off
DoR	Duration of response
ECG	Electrocardiogram
EDoR	Expected Duration of Response
EORTC QLQ-C30	European Organisation for Research and Treatment of Cancer 30-item core quality of life questionnaire

Abbreviation or special term	Explanation
EQ-5D	EuroQoL 5 dimension utility index
EQ-5D-5L	EuroQoL 5 dimension, 5 level health state utility index
FAS	Full analysis set
FWER	Family wise error rate
GSHf	Group sequential Holm fixed
HR	Hazard ratio
HRQoL	Health-related quality of life
ICU	Intensive care unit
IDMC	Independent Data Monitoring Committee
irRC	Immune-related response criteria
irRECIST 1.1	Immune-related response criteria based on RECIST 1.1
ITT	Intention to treat
iv	Intraveneous
IVRS	Interactive Voice Response System
IWRS	Interactive Web Response System
KM	Kaplan-Meier
LC13	Lung Cancer Module; 13-item self-administered questionnaire from the EORTC for lung cancer
LD	Longest diameter
MedDRA	Medical Dictionary for Regulatory Activities
mg	Milligram
MRI	Magnetic resonance imaging
NA	Not applicable
NCI	National Cancer Institute
ND	No disease, by BICR
NE	Not evaluable
NED	No evidence of disease, by Investigator assessment
NN	Non-CR/Non-PD
NSCLC	Non-small cell lung cancer
NTL	Non-target lesions
OAE	Other significant adverse event (see definition in Section 3.4.1)
ORR	Objective response rate

Abbreviation or special term	Explanation
OS	Overall survival
OS24	Proportion of patients alive at 24 months from randomisation
PD	Progressive disease
PD-L1	Programmed death ligand 1
PDx	Pharmacodynamic(s)
PFS	Progression free survival
PFS2	Time from randomisation to second progression
PGx	Pharmacogenetic(s)
PID	Percentage intended dose
PID2	Percentage intended dose (treatment through progression)
PK	Pharmacokinetic(s)
PR	Partial response
PRO	Patient reported outcome
PT	Preferred term
Q2W	Every 2 weeks
QoL	Quality of life
QTcF	QT interval (corrected for heart rate using Fridericia's correction)
RDI	Relative dose intensity
RDI2	Relative dose intensity (treatment through progression)
RECIST	Response Evaluation Criteria In Solid Tumours.
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SD	Stable disease
SBP	Systolic blood pressure
SOC	System organ class
TFST	Time to first subsequent therapy or death
TL	Target lesions
TSST	Time to second subsequent therapy or death
TTDM	Time to death or distant metastasis
ULN	Upper limit of normal
WHO	World Health Organization

AMENDMENT HISTORY

Date	Brief Description of Change
	N/A
7 April 2016	Updated based on protocol amendment 4 Abbreviations Added FWER. GSHf. irRECIST, ND, NN and TTDM Section 1.1 Changed primary RECIST 1.1 assessments to be BICR, secondary to be site investigator Added TTDM to objectives, Table 1, section 3.2.2.7 and 12.2.9 as had not been previously included Section 1.2 changed specification that radiation must be completed within 42 days (from 14 days) and randomisation may be delayed by up to 42 days (from 14). Sections 1.2, 4.1.2, 5.1 and 5.2 Updated the number of events and timelines for the interim analyses Sections 1.3 and 4.1.2 Change to alpha allocation. The alpha allocation between PFS and OS has been to (2.5%, 2.5%) from (0.5%, 4.5%) Updated Figure 1 to match revised protocol Section 2.1 Safety analysis set updated to remove criteria that post-dose data should be available Section 2.2 Violation and deviations updated based on changes to the SAP template Section 3.1 updated days the baseline tumour assessment should be taken to match the protocol. Added details about BICR assessment Sections 3.1.1, 3.1.2 and 3.1.3 Added sections on BICR RECIST 1.1, RECIST 1.1 modified for confirmation of progression and irRECIST 1.1 Section 3.2 Added information on efficacy outcomes for BICR Section 3.2.1.1 Updated to add details about how survival data will be obtained Section 3.2.1.2 Changed to BICR as primary, added details about how missing assessments will be determined Section 3.2.2.2 Changed to BICR as primary. For change in tumour size, change to BICR. Section 3.2.2.7 Time to death of distant metastasis, new endpoint added to match protocol. Section 3.3.1.2 Remove sensitivity anlaysis for HRQoL deterioration where PFS is considered an event of HRQoL deterioration Section 3.3.2 LC-13 clinically meaningful change updated to be 10 (from than 5) Section 3.3.2.1 Time to symptom deterioration. Clinically meaning symptom deterioration updated to be 10 (from 5). Sensitivity analysis where progression is considered an event removed Section 3.3.2.2 Symptom improvement rate. Clinically meaning symptom

	deterioration updated to be 10 (from 5) Section 3.4.1 Adverse Events. Changed date of collection of AEs to informed consent (from first dose) to match protocol. Added a definition of treatment emergent adverse event. AEs of special interest. Removed list of AESI Section 3.4.6 Updated definition of total exposure and calculation of duration of dose delays. Sections 3.4.7 and 4.2.1 Removed PID, RID and RID2 Section 4.1.1 Baseline. Added more detail for definition of baseline Section 4.1.2 Multiplicity. Updated as per protocol amendment. Added details of multiple testing procedure. Section 4.2 Table 8 Updated to reflect changes to protocol – BICR primary, site investigator assessment secondary. For EORTC QLQ-C30 and LC13 removed sensitivity analysis for attrition. Section 4.2.1.1 changed CI to 97.5%, from 95.5%, to match change in alpha. Added subgroup analysis for time from last dose of radiation, remove PD-L1 subgroup and added general biomarker subgroup. Clarified language for analyses of subgroups. Section 4.2.1.2 Changed to have BICR primary, site investigator secondary, including updating sensitivity analyses. Section 4.2.2.2 Changed to have BICR primary, site investigator secondary Section 4.2.2.7 new section for endpoint Time to death or distant metastasis Section 4.2.3 updated key endpoints as per protocol amendment Section 4.2.3.1 and 4.2.3.2 removed sensitivity analyses Section 4.2.4.1 Clarified that first subsequent cancer therapy doesn't include palliative radiotherapy. Removed some presentations of AEs. Removed section on long term tolerability. Section 4.2.4.2 Clarified that first subsequent cancer therapy doesn't include palliative radiotherapy. Updated tables and figures that will be presented. Section 4.2.4.4 Updated tables and figures that will be presented.
5 Dec 2016	Section 3.2.2.1 Added text to state proportion of patients alive at 12 months will be presented Section 3.2.2.6 Clarified title to include death for PFS2 Section 3.2.2.7 Clarified definition of Time to death or distant metastasis, including adding table to the Appendix Section 3.3.1 Clarified PRO baseline should be last value prior to randomisation Section 3.4.2 Added formula for creatinine clearance Table 8 and Section 4.2.1.1 Removed global interaction test for subgroup analyses, subgroups will be tested using model only including treatment

	4.2.1.1/4.2.1.2 Clarified that the confidence intervals will be 1-alpha adjusted. Section 4.2.1.1 Add details for subgroups PD-L1 and EGFR. Section 4.2.2.1 Remove HR and added p-value to the presentation of OS12 Section 4.2.2.3 Removed analysis of Duration of Response Section 4.2.2.4/5 Removed HR for APF12 and APF18 and added CI for each treatment Section 4.2.3.3 Added mixed models repeated measures for analyses of PRO endpoints Section 4.2.4.1 Updated description of death table
22 March 2017	Sections 3.2.1.2, 3.2.2.2, 3.2.2.7 Clarify criteria for censoring in the setting of no progression or death from last evaluable RECIST 1.1 assessment to last RECIST 1.1 assessment Section 4.2.1.1 removed text to state IVRS data will be used for the subgroup analyses for the stratification factors. CRF data will be used.

1. STUDY DETAILS

1.1 Study Objectives

1.1.1 Primary objective

Primary Objective:	Outcome Measure:
To assess the efficacy of MEDI4736 treatment compared with placebo in terms of OS and PFS	OS PFS using BICR assessments according to RECIST 1.1

OS Overall survival; PFS Progression free survival; RECIST Response Evaluation Criteria In Solid Tumours. BICR Blinded Independent Central Review

1.1.2 Secondary objectives

Secondary Objective:	Outcome Measure:
To further assess the efficacy of MEDI4736 compared with placebo in terms of: OS24, ORR, DoR, APF12, APF18, PFS2, and TTDM.	OS24 ORR using BICR assessments according to RECIST 1.1 DoR using BICR assessments according to RECIST 1.1 APF12 and APF18 using BICR assessment according to RECIST 1.1 PFS2 as defined by local standard clinical practice TTDM using BICR assessments according to RECIST 1.1
To assess the safety and tolerability profile of MEDI4736 compared with placebo	AEs, physical examinations, vital signs, pulse, electrocardiograms, and laboratory findings including clinical chemistry, haematology and urinalysis
To assess the PK of MEDI4736	Concentration of MEDI4736 in blood and non-compartmental PK parameters (such as peak concentration and trough, as data allow) (sparse sampling)
To investigate the immunogenicity of MEDI4736	ADA (confirmatory results: positive or negative; titres [ADA neutralising antibodies will also be assessed])

Secondary Objective:	Outcome Measure:
To assess symptoms and health-related quality of life in patients treated with MEDI4736 compared with placebo using EORTC QLQ-C30 v3 and LC13	EORTC QLQ-C30: Time to symptom deterioration (fatigue, pain, nausea/vomiting, dyspnoea, loss of appetite, insomnia, constipation, and diarrhoea). Time to QoL/function deterioration (physical function; role function; emotional function; cognitive function; social function and global health status/QoL) LC13: Time to symptom deterioration (dyspnoea, cough, haemoptysis, chest pain, arm/shoulder pain, other pain) Changes in World Health Organization Performance Status will also be assessed

ADA Anti-drug antibody; AE Adverse event; APF12 Proportion of patients alive and progression free at 12 months from randomisation; APF18 Proportion of patients alive and progression free at 18 months from randomisation; BICR Blinded Independent Central Review; DoR Duration of response; EORTC QLQ-C30 European Organisation for Research and Treatment of Cancer 30-item core quality of life questionnaire; LC13 Lung Cancer Module; ORR Objective response rate; OS24 Proportion of patients alive at 24 months from randomisation; PFS2 Time from randomisation to second progression; PK Pharmacokinetic(s); QoL Quality of Life; RECIST Response Evaluation Criteria In Solid Tumours; TTDM Time to death or distant metastasis.

1.1.3 Exploratory objectives

Exploratory Objective:	Outcome Measure:
To explore irRC (irRECIST 1.1) criteria as an assessment methodology for clinical benefit of MEDI4736 compared with placebo by BICR	PFS, and ORR using BICR assessments according to irRC (irRECIST 1.1)
To investigate the relationship between MEDI4736 PK exposure and clinical outcomes, efficacy, AEs and/or safety parameters, if deemed appropriate	A graphical and/or a data modelling approach will be used to analyse MEDI4736 PK exposure and the relationship with clinical outcomes, efficacy, AEs and/or safety parameters, as deemed appropriate
To describe and evaluate resource use associated with MEDI4736 treatment and underlying disease	Health resource utilisation measures including hospitalization, outpatient visits, or emergency department visits
To explore the impact of treatment and disease state on health state utility using the EQ-5D-5L	The EQ-5D-5L health state utility index will be used to derive health state utility based on patient reported data
To investigate the relationship between a patient's PD-L1 expression and spatial distribution within the tumour microenvironment and efficacy outcomes with MEDI4736	Tumoural expression of PD-L1 and spatial distribution within the tumour microenvironment relative to efficacy outcomes (OS, PFS and ORR)
To collect blood and tissue samples for analysis of peripheral and tumoural biomarkers	Biomarker analysis of blood and tissue to assess exploratory markers which may include but is not limited to: immune cell gene expression profiles within the peripheral and tumoural compartments, the presence of IFN-γ tumour necrosis factor-α, IL-2, IL-6, IL-10, IL-8, and IL-12 as well as antibodies against tumour, self, or viral antigens, expression of PD-L1 and the number and phenotype of immune cells such as T-cells

Exploratory Objective:	Outcome Measure:
To explore the relationship(s) between a patient's biomarker status and MEDI4736 PK exposure and clinical outcomes before and after treatment	Biomarker status before and after treatment and MEDI4736 PK exposure and relationship with clinical outcomes, efficacy, AEs and/or safety parameters, as deemed appropriate
To explore potential biomarkers in residual biological samples (eg, tumour, plasma and/or serum), which may influence the progression of cancer (and associated clinical characteristics) and/or prospectively identify patients likely to respond to MEDI4736 treatment	Correlation of biomarkers with response to MEDI4736 treatment and/or the progression of cancer
CCI [redacted]	CCI [redacted]

AE Adverse event; BICR Blinded Independent Central Review; DoR Duration of response; EQ-5D-5L EuroQoL 5 dimension, 5 level health state utility index; IFN Interferon; IL Interleukin; irRC Immune-related response criteria; ORR Objective response rate; PD-L1 Programmed death ligand 1; PDx Pharmacodynamic(s); PFS Progression free survival; PK Pharmacokinetic(s); T-cell T lymphocyte.

With regards to biomarkers, Programmed death ligand 1 (PD-L1) expression determined by immunohistochemistry will be reported in the Clinical Study Report (CSR). Other exploratory biomarker and Pharmacogenetic (PGx) research will be reported outside the CSR. Data relating to the exploratory objectives of patient reported outcome (PRO) and Health Related Quality of Life (HRQoL) will be reported in the CSR.

1.2 Study Design

This study is a Phase III, randomised, double-blind, placebo-controlled, multi-centre study assessing the efficacy and safety of MEDI4736 compared with placebo as sequential therapy in male and female patients with locally advanced, unresectable non-small cell lung cancer (NSCLC) (Stage III) who have not progressed following definitive, platinum-based, concurrent chemoradiation therapy.

Approximately 880 patients with locally advanced, unresectable NSCLC (Stage III) will be recruited and 702 patients randomised at 260 to 330 sites in Australia, Asia, Europe, North and South America and South Africa. These patients will be in complete response (CR), partial response (PR), or have stable disease (SD) following definitive, platinum-based, concurrent chemoradiation therapy.

Patients will be randomised in a 2:1 ratio (MEDI4736 to placebo) to 1 of 2 arms:

- MEDI4736 (10 mg/kg Q2W iv for up to 12 months).
- Placebo (matching placebo for infusion Q2W iv for up to 12 months).

Randomisation will be stratified by: age at randomisation (<65 vs ≥65 years of age), sex (male vs female), and smoking history (smoker vs non-smoker). Patients must not have progressed following definitive, platinum-based, concurrent chemoradiation therapy; radiation therapy must be completed within 1 to 42 days prior to randomisation in the study (the last dose of radiation therapy is defined as the day of the last radiation treatment session). For patients who are recovering from toxicities associated with prior treatment, randomisation may be delayed by up to 42days from the end of the chemoradiation therapy. For those patients randomised to the placebo arm no cross-in to the MEDI4736 arm is permitted and similarly those patients randomised to the MEDI4736 arm no cross-in to the placebo arm is permitted.

Administration of study drug will commence on Day 1 following randomisation to MEDI4736 or placebo after confirmation of eligibility and will continue on a Q2W schedule for a maximum duration of 12 months. Study drug should be discontinued prior to 12 months if there is confirmed PD (unless the investigator considers the patient continues to receive benefit from the study drug), initiation of alternative cancer therapy, unacceptable toxicity, withdrawal of consent, or other reasons to discontinue study drug occur.

Tumour assessments using computed tomography (CT)/magnetic resonance imaging (MRI) will be performed at the times specified in Table 1, Table 2 and Table 3 of the Clinical Study Protocol (CSP). Response Evaluation Criteria In Solid Tumours 1.1 (RECIST 1.1) measurements (using Blinded Independent Central Review [BICR] data) will be used to derive the co-primary variable of PFS and secondary variables of objective response rate (ORR), duration of response (DoR), proportion of patients alive and progression free at 12 months (APF12), and proportion of patients alive and progression free at 18 months (APF18). Categorisation of objective tumour response assessment will be based on RECIST 1.1: complete response (CR), partial response (PR), stable disease (SD) and progressive disease (PD). Patients with no evidence of disease at follow-up in the absence of new lesions will be assigned a response of no evidence of disease (ND). Once a patient has had objective progression recorded and has discontinued study drug, the patient will be followed up for survival status every 2 months until death, withdrawal of consent or the end of the study. Patients will also be assessed every 12 weeks for a second progression defined according to local standard clinical practice and may involve any of: objective radiological progression, symptomatic progression or death.

There will be 4 data cut-off (DCO) time points in the study. The first analysis DCO will occur when it is expected that 367 PFS events have occurred (52% maturity, approximately 30 months after the first patient is randomised) and be conducted on the PFS endpoint. The second analysis DCO will occur at the time of the primary PFS analysis when it is expected that 458 PFS events have occurred (65% maturity, approximately 36 months after the first patient is randomised) and the first OS interim analysis will be conducted at the same time (with approximately 285 events, 41% maturity). The third analysis DCO will occur at the time of the second OS interim analysis when it is expected that 393 OS events have occurred (56% maturity, approximately 47 months after the first patient is randomised). The fourth analysis DCO will occur at the time of the primary OS analysis when it is expected that 491

OS events have occurred (70% maturity, at approximately 62 months). See Section 5 for further details.

Sensitivity analyses will be performed on PFS and ORR based on data from site investigator assessment based upon RECIST 1.1 and BICR data per RECIST 1.1 modified for confirmation of progression. Exploratory analysis of PFS and ORR will also be performed for data obtained from the BICR using immune-related response criteria (irRECIST 1.1). See Section 6.3 and Appendix F of the CSP for further information regarding RECIST 1.1 tumour assessments in this study.

The study flow chart is presented in Figure 1.

The schedule of study procedures at Screening and during the Treatment Period is presented in Table 1 of the CSP. The schedule of study procedures during follow-up for patients who have completed the Treatment Period and achieved disease control (until confirmed PD) and patients who have discontinued study drug due to toxicity or a reason other than confirmed PD is presented in Table 2 of the CSP. The schedule of study procedures during follow-up for patients who have discontinued study drug due to confirmed PD is presented in Table 3 of the CSP.

Guidelines for the management of toxicities are described in Section 5.5.3 and Appendix H of the CSP.

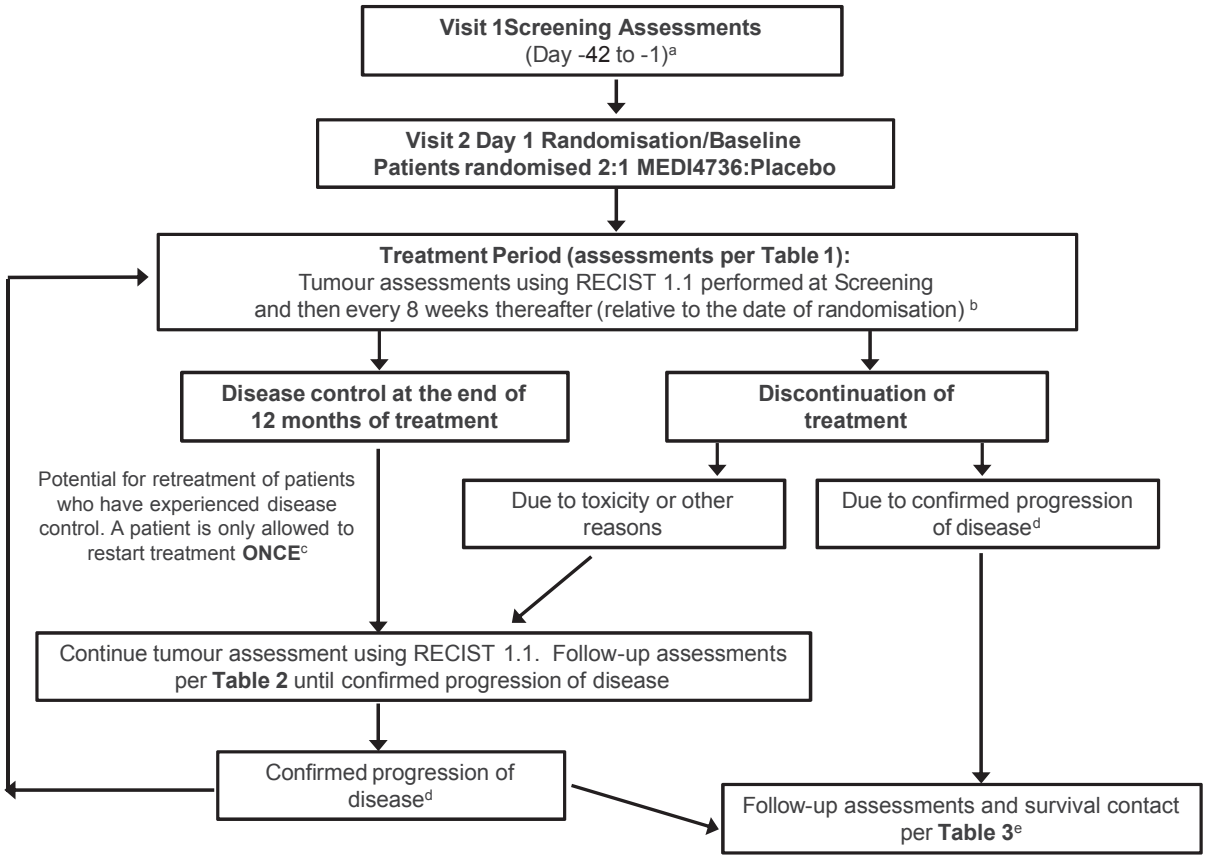
Details of the PGx component of the study (relating to DNA) are provided in Appendix D of the CSP.

Independent Data Monitoring Committee

An Independent Data Monitoring Committee (IDMC) will be convened and will meet approximately 3 months after the study has started, or once 75 patients have been randomised whichever occurs first, to review safety assessments and make recommendations to continue, amend, or stop the study based on safety findings. The committee will then meet again 3 months later and then at least every 6 months thereafter. All patients who receive a dose of study drug will be evaluated for safety and tolerability. Enrolment will continue unless there is an unexpected safety concern. The study may be adjusted or suspended depending on the IDMC review outcome. The first three efficacy analyses will also be assessed by the IDMC.

Details on the IDMC are provided in Section 12.4 of the CSP and full details of the IDMC procedures and processes can be found in the IDMC Charter.

Figure 1 Study flow chart



a Screening assessments can be performed in a step-wise process. The baseline tumour assessment is part of the screening procedures and should be performed within 0 to 42 days after the end of chemoradiation therapy and before the start of study drug. For patients who are recovering from toxicities associated with prior treatment, randomisation may be delayed by up to 42 days from the end of the chemoradiation therapy.

b Disease progression needs to be confirmed, the confirmatory scan should occur preferably at the next scheduled visit and no earlier than 4 weeks after the initial assessment of PD in the absence of clinically significant deterioration. Administration of study drug will continue between the initial assessment of progression and confirmation for progression. For all patients who are treated through progression, the Investigator should ensure patients do not have any significant, unacceptable or irreversible toxicities that indicate continuing treatment will not further benefit the patient. The patient must continue to meet those inclusion and exclusion criteria that are relevant to treatment through disease progression and retreatment as specified in Section 4.3 of the CSP. Patients with rapid tumour progression or with symptomatic progression that requires urgent medical intervention (eg, central nervous system metastasis, respiratory failure due to tumour compression, spinal cord compression) will not be eligible to continue to receive study drug.

c Patients who achieve and maintain disease control (ie, CR, PR, no evidence of disease, or SD) through to the end of the 12-month treatment period, may restart study drug upon evidence of PD (according to RECIST 1.1), with or without confirmation, during follow-up. Before restarting study drug, the Investigator should ensure patients still meet all of the inclusion criteria and none of the exclusion criteria for this study including re-consenting to treatment. To restart study drug the patient must not have received an intervening systemic anti-cancer therapy post study drug discontinuation. Patients should have a baseline tumour assessment within 28 days of restarting study drug, all further scans should occur every 8 weeks (relative to the date of restarting study drug) (maximum of 12 months of further treatment).

d Patients with confirmed PD that continue to receive study drug at the discretion of the Investigator (following consultation with the sponsor) can receive study drug for a maximum of 12 months. For all patients who are treated through progression, the Investigator should ensure patients do not have any significant, unacceptable or irreversible toxicities that indicate continuing treatment will not further benefit the patient. The patient must continue to meet those inclusion and exclusion criteria that are relevant to treatment through disease progression and retreatment as specified in Section 4.3 of the protocol. The same exceptions as noted in footnote b apply. Patients will follow the assessments in Table 1 of the CSP including tumour assessments every 8 weeks (relative to the date of randomisation) until study drug is stopped. Study drug should be discontinued if there is confirmed progression of disease (PD) following a previous response (PR or CR) to study drug.

e Patients with confirmed PD that discontinue study drug, should have scans conducted according to local standard clinical practice (see Section 6.2.2.2 of the CSP) that are submitted for BICR until the patient commences a new treatment (these scans are optional).

Note: The DCO for the analysis of overall survival will take place once 491 death events have occurred (estimated to be approximately 62 months after randomisation). At this time point, patients who are receiving study drug can either choose to discontinue from the study or where the Investigator believes patients are gaining clinical benefit, patients may continue to receive study drug. For patients who do continue to receive study drug beyond the time of the final DCO, Investigators will only continue to report all SAEs to Quintiles Patient Safety until 90 days after study drug is discontinued and any SAE or non-serious AE that is ongoing at the time of this DCO must be followed up to resolution unless the event is considered by the Investigator to be unlikely to resolve, or the patient is lost to follow-up.

AE Adverse event; BICR Blinded Independent Central Review; CR Complete response; DCO Data cut-off; PD Progression of disease; PR Partial response; RECIST Response Evaluation Criteria In Solid Tumours; SAE Serious adverse event; SD Stable disease.

1.3 Number of Subjects

The sample size for this study was selected to be consistent with the research hypothesis as described in Section 1.2 of the CSP.

The two co-primary endpoints of this study are OS and PFS. To control for type-I error, a significance level of 2.5% will be used for analysis of OS and a significance level of 2.5% will be used for analysis of PFS. The study will be considered positive (a success) if either the PFS analysis results and/or the OS analysis results are statistically significant.

Approximately 702 patients will be randomised 2:1 to obtain 491 death events in the ITT population (70% maturity). The primary PFS analysis DCO will occur when it is expected that 458 PFS events have occurred (65% maturity). If the true PFS HR is 0.67, the study will provide at least 95% power to demonstrate a statistically significant difference for PFS with a 2-sided significance level of 2.5% in the ITT population; this translates to a 5-month benefit in median PFS over 10 months on placebo if PFS is exponentially distributed. The smallest treatment difference that would be statistically significant is a HR of 0.8. A recruitment period of approximately 22 months and a follow-up period of 14 months are expected for the PFS endpoint. Therefore it is anticipated that this PFS analysis could be performed approximately 36 months after the first patient has been randomised.

The primary OS analysis DCO will occur when it is expected that 491 OS events have occurred (70% maturity). If the true OS HR is 0.73, this number of death events will provide at least 85% power to demonstrate a statistically significant difference for OS, assuming a 2.5% 2-sided significance level in the ITT population; this translates to an 8-month benefit in median OS over 22 months on placebo if OS is exponentially distributed (Butts et al 2014). The smallest treatment difference that would be statistically significant is a HR of 0.81. A recruitment period of approximately 22 months and a follow-up period of 40 months are expected for the OS endpoint. Therefore it is anticipated that this OS analysis could be performed at approximately 62 months after the first patient has been randomised.

2. ANALYSIS SETS

2.1 Definition of Analysis Sets

Four main analysis sets are defined for this study.

Full analysis set (FAS) (Intention to treat (ITT)):

The primary statistical analysis of the efficacy of MEDI4736 vs placebo will include all randomised patients and will compare the treatment arms on the basis of randomised treatment, regardless of the treatment actually received. Patients who were randomised but did not subsequently go on to receive study drug are included in the ITT population. Note, this is also known as the FAS. Therefore, all efficacy and HRQoL data will be analysed using the FAS on an ITT basis.

Safety analysis set:

All patients who received at least one dose of randomised study medication, MEDI4736 or placebo, (regardless of whether that was the randomised therapy intended or indeed whether, in rare cases, they received therapy without being randomised) will be included in the safety population. Throughout the safety sections, erroneously treated patients (e.g., those randomised to Treatment A but actually given Treatment B) will be accounted for by the actual treatment administered. As long as a patient received any dose of MEDI4736, the patient would be counted under MEDI4736 treatment group.

When assessing safety and tolerability, summaries will be produced based on the safety analysis set.

PK analysis set:

All patients who receive at least 1 dose of MEDI4736 per the protocol, for whom any post-dose data are available and do not violate or deviate from the protocol in ways that would significantly affect the PK analyses will be included in the PK analysis set. The population will be defined by the Study Team Physician, Pharmacokineticist and Statistician prior to any analyses being performed.

Table 1 gives a summary of outcome variables and analysis populations.

Table 1 Summary of outcome variables and analysis populations

Outcome variable	Populations
Efficacy Data	
OS (co-primary)	FAS (ITT)
PFS (co-primary)	FAS (ITT)
OS24, APF12, APF18, PFS2, DoR, TTDM, TFST, PRO endpoints, World Health Organisation performance status	FAS (ITT)
ORR	FAS (ITT)
Study Population/Demography Data	
Demography characteristics (e.g. age, sex etc.)	FAS (ITT)
Baseline, disease characteristics	FAS (ITT)
Analysis populations	FAS (ITT)
Important deviations	FAS (ITT)
Medical/Surgical history	FAS (ITT)
Previous anti-cancer therapy	FAS (ITT)
Concomitant medications/procedures	FAS (ITT)
Subsequent anti-cancer therapy	FAS (ITT)
World Health Organization performance status	FAS (ITT)

Table 1 Summary of outcome variables and analysis populations

Outcome variable	Populations
PK Data	
PK data	PK
Immunogenicity	
Immunogenicity data	Safety
Safety Data	
Exposure	Safety
Adverse events	Safety
Laboratory measurements	Safety
Vital Signs	Safety
ECGs	Safety

APF12 Proportion of patients alive and progression free at 12 months from randomisation; APF18 Proportion of patients alive and progression free at 18 months from randomisation; DoR Duration of response;; ITT Intent-to-Treat; ORR Objective response rate; OS Overall survival; OS24 Proportion of patients alive at 24 months from randomisation; PFS Progression free survival; PFS2 Time from randomisation to second progression; PK Pharmacokinetic; PRO Patient reported outcomes; TTDM Time to death or distant metastasis, TFST Time to first subsequent therapy or death.

2.2 Protocol Deviations

The important protocol deviations will be listed and summarised by randomised treatment group. Deviation 1 below will lead to exclusion from the Safety analysis set. This means that this will not be shown on the relevant summary but can still be identified on the listings. None of the other deviations will lead to any patient being excluded from any of the analysis sets described in Section 2.1 (with the exception of the PK analysis set, if the deviation is considered to impact upon PK). If the deviations are serious enough to have the potential to impact the primary analysis, sensitivity analyses may be performed. The need for such a sensitivity analysis will be determined following review of the protocol deviations ahead of database lock (DBL) and will be documented prior to the primary analysis being conducted. Eligibility criteria deviations are deviations from the protocol inclusion and exclusion criteria. Post-entry deviations are deviations from the protocol that occurred after the patient was assigned to the study treatment.

The following general categories will be considered important deviations and be listed and discussed in the CSR as appropriate for the study:

- Deviation 1: Patients randomised but who did not receive MEDI4736/matching placebo
- Deviation 2: Patients who deviate from the following key entry criteria in CSP Amendment:

  a) Inclusion criteria: 3, 4, 5, 8

b) Exclusion criteria: 5, 6, 7, 8, 10, 12, 13, 14, 17

- Deviation 3: Baseline RECIST scan > 42 days before randomisation.
- Deviation 4: No baseline RECIST 1.1 assessment on or before date of randomisation.
- Deviation 5: Received prohibited concomitant systemic anti-cancer agents.

  Refer to the CSP section 5.6 for those medications that are detailed as being 'excluded' from permitted use during the study. This will be used as a guiding principle for the physician review of all medications prior to DBL.

- Deviation 6: Patients randomised who received treatment other than that to which they were randomised to.

The categorisation of these as important deviations is not automatic and will depend on duration and the perceived effect on efficacy.

In addition to the programmatic determination of the deviations above, monitoring notes or summaries will be reviewed to determine any important post entry deviations that are not identifiable via programming, and to check that those identified via programming are correctly classified. The final classification will be made at the blind data review meeting (BDRM) prior to DBL or DCO. Decisions made at the BDRM will be documented and approved by AstraZeneca prior to analysis.

Errors in treatment dispensing, in addition to incorrect stratifications, will also be investigated more closely. This may occur when a patient is not randomised or treated according to the randomisation schedule on at least one occasion. It is envisaged that there will be 3 sub categories of this within the important deviations summary:

- Patients who receive no treatment whatsoever for a period of time due to errors in dispensing of medication. Note, this is not due to tolerability issues where patients may stop taking drug.
- The patient receives a treatment pack with a different code to their randomisation code and the treatment differs from the randomized treatment.
- The patient receives a treatment pack with a different code to their randomisation code. However, the actual treatment may still match the randomised treatment. For example, a patient is given randomisation code 0001, which according to the randomisation schedule is MEDI4736. However, at the randomisation visit they are given treatment pack 0003, but this still contains MEDI4736.

Patients who receive the wrong treatment at any time will be included in the safety analysis set as described in Section 2.1. During the study, decisions on how to handle errors in treatment dispensing (with regards to continuation/discontinuation of study treatment or, if

applicable, analytically) will be made on an individual basis with written instruction from the study team leader and/or statistician.

3. PRIMARY, SECONDARY AND EXPLORATORY VARIABLES

3.1 Derivation of RECIST Visit Responses

For all patients, the RECIST version 1.1 (see further Appendix F of the CSP and Independent Review Charter) tumour response data will be used to determine each patient's visit response. It will also be used to determine if and when a patient has progressed and also their best objective response.

The baseline tumour assessment is part of the screening procedures and should be performed within 0 to 42 days after the end of chemoradiation therapy and no more than 28 days before the start of study drug. For patients who are recovering from toxicities associated with prior treatment, randomisation may be delayed by up to 42 days from the end of the chemoradiation therapy.

Efficacy for all patients will be assessed by objective tumour assessments every 8 weeks (relative to the date of randomisation) for the first 12 months, then every 12 weeks thereafter, until confirmed objective disease progression as defined by RECIST 1.1 (irrespective of the reason for stopping study drug and/or subsequent therapy). For patients who discontinue study drug due to toxicity or a reason other than confirmed PD, objective tumour assessments should be continued every 8 weeks for 48 weeks (relative to the date of randomisation) then every 12 weeks thereafter until confirmed objective disease progression. If an unscheduled assessment is performed, and the patient has not progressed, every attempt should be made to perform the subsequent assessments at their scheduled visits.

Disease progression requires confirmation, the confirmatory scan should occur preferably at the next scheduled visit and no earlier than 4 weeks after the initial assessment of PD in the absence of clinically significant deterioration. Study treatment will continue between the initial assessment of progression and confirmation for progression. For all patients who are treated through progression and patients who achieve disease control (i.e., CR, PR or SD) through to the end of the 12 month treatment period and restart study treatment upon evidence of PD (according to RECIST 1.1), with or without confirmation, during follow-up, the investigator should ensure patients do not have any significant, unacceptable or irreversible toxicities that indicate continuing treatment will not further benefit the patient, and that the patient still meets all of the inclusion criteria and none of the exclusion criteria for this study including re-consenting to continue or restart treatment. To restart study treatment the patient must not have received an intervening systemic anti-cancer therapy post study drug discontinuation. Study drug should be discontinued if there is confirmed PD following a previous response (PR or CR) to study drug.

Patients with rapid tumour progression or with symptomatic progression that requires urgent medical intervention (eg, central nervous system metastasis, respiratory failure due to tumour compression, spinal cord compression) will not be eligible to continue to receive study drug.

Progression would be considered confirmed if the following criteria are met:

- ≥20% increase in the sum diameters of target lesions (TLs) compared with the nadir at 2 consecutive visits with an absolute increase of at least 5 mm

  The assessment of progression of ≥20% increase in the sum diameters of TLs compared with the nadir is at the first progression time point relative to the nadir (the smallest sum of diameters and this may be at baseline or subsequent follow-up visit). The confirmed scan confirms the persistence of the ≥20% increase relative to the nadir.

- and/or significant progression (worsening) of non target lesions (NTLs) or new lesions at the confirmatory PD time point compared with the first time point where progression of NTLs or new lesions identified

- and/or additional new unequivocal lesions at the confirmatory PD time point compared with the first time point new lesions identified.

BICR according to RECIST 1.1 will be regarded as primary in terms of the efficacy analyses. Sensitivity analyses will be performed using investigator assessment data and BICR using RECIST 1.1 modified for confirmation of progression. Additionally, the BICR will also be performed using irRECIST 1.1(Wolchok et al 2009, Nishino et al 2013)Nishino et al 2013, see also Section 3.1.3) for exploratory purposes.

From the investigator's review of the imaging scans, the RECIST tumour response data will be used to determine each patient's visit response according to RECIST version 1.1. At each visit, patients will be programmatically assigned a RECIST 1.1 visit response of CR, PR, SD or PD, using the information from target lesions (TLs), non-target lesions (NTLs) and new lesions and depending on the status of their disease compared with baseline and previous assessments. If a patient has had a tumour assessment which cannot be evaluated then the patient will be assigned a visit response of not evaluable (NE) (unless there is evidence of progression in which case the response will be assigned as PD).

RECIST outcomes (i.e., PFS and ORR etc) will be calculated programmatically for the BICR and site investigator data from the overall visit responses.

3.1.1 Blinded Independent Central Review (BICR) Assessment Using RECIST 1.1

A planned BICR will be carried out using RECIST version 1.1. All radiological scans for all patients (including those at unscheduled visits, or outside visit windows) will be collected on an ongoing basis and sent to an AstraZeneca appointed Contract Research Organisation (CRO) for central analysis. Information prior radiotherapy from the CRF will also be

provided to the BICR to delineate the anatomical region of previous definitive radiotherapy, to confirm that there are no lesions at baseline outside the prior radiotherapy region, to allow the selection of appropriate target lesions, if any at baseline, and to designate new lesions at follow-up that are outside the region of prior radiotherapy as 'distant' metastases for derivation of TTDM. The imaging scans will be reviewed by 2 independent radiologists using RECIST 1.1 and will be adjudicated, if required (i.e. two primary reviewers review the scans and adjudication is performed by a separate reviewer in case of a disagreement). The adjudicator must choose the assessments of one of the two primary reviewers. For each patient, the BICR will define the overall visit response (ie, the response obtained overall at each visit by assessing TLs, NTLs and new lesions) data and no programmatic derivation of visit response is necessary. Possible overall visit responses include: for patients with TLs at baseline: CR, PR, SD, PD, not evaluable [NE]; for patients with NTLs at baseline only: CR, non-CR/non-PD (NN), PD or NE; for patients with no disease identified at baseline: PD, no evidence of disease [ND], NE. If a patient has had a tumour assessment that cannot be evaluated then the patient will be assigned a visit response of NE (unless there is evidence of progression in which case the response will be assigned as PD). RECIST assessments/scans contributing towards a particular visit may be performed on different dates and for the central review the date of progression for each reviewer will be provided based on the earliest of the scan dates of the component that triggered the progression.

If adjudication is performed, the records from the primary reviewer with which the adjudicator agreed will be used for data analysis (note in the case of more than one review period, the latest adjudicator decision will be used). In the absence of adjudication, the records for all visits for a single reviewer will be used. The reviewer selected in the absence of adjudication will be the reviewer who read the baseline scan first. The records from the single selected reviewer will be used to derive all BICR RECIST information including dates of progression, visit response, censoring and changes in target lesion dimensions. Endpoints (of PFS, ORR and DoR) will be derived programmatically from this information.

Results of this independent review will not be communicated to Investigators and the management of patients will be based solely upon the results of the RECIST 1.1 assessment conducted by the Investigator.

The BICR of all patients up to each DCO will be completed before the data-base lock for the interim analysis. Additionally the BICR of all patients will be performed for the final database lock for PFS which will cover all of the scans up to that point.

Further details of the BICR will be documented in the Independent Review Charter.

3.1.2 BICR assessment using RECIST 1.1 Modified for Confirmation of Progression

The BICR of all scans used in the assessment of tumours will be assessed using RECIST 1.1 modified for confirmation of progression as detailed in Sections 3.1 and 3.1.1. This means that the visit response of PD must be confirmed by another visit response of PD at the next

scheduled visit and no earlier than 4 weeks after the initial assessment of PD in the absence of clinical deterioration. Confirmation of progression needs to be programmatically derived.

3.1.3 BICR assessment using irRECIST 1.1

The criteria that built the foundation for irRC were the WHO criteria which existed before the publication of RECIST 1.1. The irRC are, therefore, like WHO criteria, based on bi-dimensional measurements (Wolchok et al 2009). More recently, RECIST 1.1 has become the gold standard for disease status determination for study subjects with solid tumours. To be more in line with RECIST 1.1 criteria, irRC has been adjusted so that uni-dimensional measurements are used (referred to as irRECIST 1.1). This modification of irRC 2009 criteria (irRECIST 1.1) is suggested based on the results of Nishino et al 2013, who found that irRC using the unidimensional measurements provide highly concordant response assessment compared with the bi-dimensional irRC, with less measurement variability. Furthermore, this enables a consistency with RECIST 1.1 which enables a clearer contrast of the results from the two methods than would otherwise be possible.

The definitions of CR, PR, SD and PD according to irRECIST 1.1 are outlined clearly in the Independent Review Charter (referred to as irCR, irPR, irSD/irNN and irPD in that document), but a brief description of the methodology is given here.

In irRECIST 1.1 the presence of new lesions will not automatically trigger a declaration of Progressive Disease, but instead new lesions can be measured and these measurements will be added to the sum of diameters of the target lesions. Based on the sum of these measurements and % calculations thereof, the target lesion + new measured lesion response assessment will be derived. The overall response assessment (irCR, irPR, irSD/irNN, irPD, irNE or irND) will be obtained at the BICR and confirmation of PD is required. The imaging scans will be reviewed by 2 independent radiologists and will be adjudicated, if required.

Where stated, the endpoints in Section 3.2 will also be derived using these assessment methods for supportive purposes.

3.1.4 Site investigator assessment using RECIST 1.1: target lesions

Measurable disease is defined as having at least one measurable lesion which is ≥ 10 mm in the longest diameter (except lymph nodes which must have short axis ≥ 15 mm) with CT or MRI and which is suitable for accurate repeated measurements.

A patient can have a maximum of 5 measurable lesions recorded at baseline with a maximum of 2 lesions per organ (representative of all lesions involved suitable for accurate repeated measurement) and these are referred to as TLs. If more than one baseline scan is recorded then measurements from the one that is closest to randomisation will be used to define the baseline sum of TLs. It may be the case that, on occasion, the largest lesion does not lend itself to reproducible measurement in which circumstance the next largest lesion, which can be measured reproducibly, should be selected.

All other lesions (or sites of disease) not recorded as TL should be identified as NTL at baseline. Measurements are not required for these lesions, but their status should be followed at subsequent visits.

Note: For patients who do not have measurable disease at entry (i.e. no TLs) but have non-measurable disease, evaluation of overall visit response at each visit will be based on the overall NTL assessment and the absence/presence of new lesions (see Section 3.1.6 for further details). If a patient does not have measurable disease at baseline then the TL visit response will be not applicable (NA).

For patients with NED at baseline (i.e. no TLs and no NTLs), evaluation of overall visit response at each visit will be based on absence/presence of new lesions. If no TLs and no NTLs are recorded at a visit, both the TL and NTL visit response will be recorded as NA and the overall visit response will be NED if there is no new lesion.

If a patient has had a tumour assessment which cannot be evaluated then the patient will be assigned a visit response of not evaluable (NE) (unless there is evidence of progression in which case the response will be assigned as PD).

Table 2 TL Visit Responses

Visit Responses	Description
Complete Response (CR)	Disappearance of all TLs since baseline. Any pathological lymph nodes selected as TLs must have a reduction in short axis to <10 mm
Partial Response (PR)	At least a 30% decrease in the sum of diameters of TL, taking as reference the baseline sum of diameters as long as criteria for PD are not met.
Progressive Disease (PD)	At least a 20% increase in the sum of diameters of TLs taking as reference the smallest sum on study (this includes the baseline sum if that is the smallest on study). In addition to the relative increase of 20%, the sum must also indicate an absolute increase of at least 5 mm
Stable Disease (SD)	Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD
Not Evaluable (NE)	Only relevant if any of the TLs were not assessed or not evaluable or had a lesion intervention at this visit. Note: If the sum of diameters meets the progressive disease criteria, progressive disease overrides not evaluable as a TL response
Not Applicable (NA)	No TLs are recorded at baseline

Rounding of TL data

For calculation of PD and PR for TLs, percentage changes from baseline and previous minimum should be rounded to 1 decimal place before assigning a TL response. For example 19.95% should be rounded to 20.0% but 19.94% should be rounded to 19.9%

Missing TL data

For a visit to be evaluable, all TL measurements should be recorded. However, a visit response of PD should still be assigned if any of the following occurred:

- A new lesion is recorded.
- A NTL visit response of PD is recorded.
- The sum of TLs is sufficiently increased to result in a 20% increase, and an absolute increase of ≥ 5mm, from nadir even assuming the non-recorded TLs have disappeared.

Note: the nadir can only be taken from assessments where all the TLs had a lesion diameter recorded.

Lymph nodes

For lymph nodes, if the size reduces to < 10 mm then these are considered non-pathological. However a size will still be given and this size should still be used to determine the TL visit response as normal. In the special case where all lymph nodes are < 10 mm and all other TLs are 0 mm then although the sum may be >0 mm the calculation of TL response should be over-written as a CR.

TL visit responses subsequent to CR

A CR can only be followed by CR, PD or NE. If a CR has occurred then the following rules at the subsequent visits must be applied:

- Step 1: If all lesions meet the CR criteria (i.e. 0 mm or < 10 mm for lymph nodes) then response will be set to CR irrespective of whether the criteria for PD of TL is also met i.e. if a lymph node longest diameter (LD) increases by 20% but remains < 10 mm.
- Step 2: If some lesion measurements are missing but all other lesions meet the CR criteria (i.e. 0 mm or < 10 mm for lymph nodes) then response will be set to NE irrespective of whether when referencing the sum of TL diameters the criteria for PD is also met.
- Step 3: If not all lesions meet the CR criteria and the sum of lesions meets the criteria for PD then response will be set to PD
- Step 4: If after steps 1 – 3 a response can still not be determined the response will be set to remain as CR

TL too big to measure

If a TL becomes too big to measure this should be indicated in the database and a size ('x') above which it cannot be accurately measured should be recorded. If using a value of x in the calculation of TL response would not give an overall visit response of PD, then this will be flagged and reviewed by the study team blinded to treatment assignment. It is expected that a visit response of PD will remain in the vast majority of cases.

TL too small to measure

If a TL becomes too small to measure a value of 5 mm will be entered into the database and used in TL calculations, unless the radiologist has indicated and entered a smaller value that can be reliably measured. If a TL response of PD results then this will be reviewed by the study team blinded to treatment assignment.

Irradiated lesions/lesion intervention during the study

According to the study design, it is possible that patients have some residual disease that has been irradiated prior to their study entry. For this reason, previously irradiated lesions may be selected as TLs.

Any TL (including lymph nodes), which has had intervention during the study (for example, irradiation / palliative surgery / embolisation), should be handled in the following way and once a lesion has had intervention then it should be treated as having intervention for the remainder of the study noting that an intervention will most likely shrink the size of tumours:

- Step 1: the diameters of the TLs (including the lesions that have had intervention) will be summed and the calculation will be performed in the usual manner. If the visit response is PD this will remain as a valid response category.

- Step 2: If there was no evidence of progression after step 1, treat the lesion diameter (for those lesions with intervention) as missing and scale up as described below, as long as there remain ≤ 1/3 of the TLs with interventions. If the scaling results in a visit response of PD then the patient would be assigned a TL response of PD.

- Step 3: If after both steps PD has not been assigned, then if appropriate, a scaled sum of diameters will be calculated (as long as ≤ 1/3 of the TLs with interventions), and PR or SD then assigned as the visit response. Patients with intervention are evaluable for CR as long as all non-intervened lesions are 0 (or <10mm for lymph nodes) and the lesions that have been subjected to intervention also has a value of 0 recorded.

At subsequent visits the above steps will be repeated to determine the TL and overall visit response. When calculating the previous minimum, lesions with intervention should be treated as missing and scaled up where appropriate (as per step 2 above).

Scaling (applicable only for irradiated lesions/lesion intervention during the study)

If > 1/3 of TL measurements are treated as missing (because of intervention) then TL response will be NE, unless the sum of diameters of non-missing TL would result in PD (i.e. if using a value of 0 for missing lesions, the sum of diameters has still increased by > 20% or more compared to nadir and the sum of TLs has increased by 5mm from nadir).

If ≤ 1/3 of the TL measurements are treated as missing (because of intervention) then the results will be scaled up (based on the sizes at the nadir visit to give an estimated sum of diameters and this will be used in calculations; this is equivalent to comparing the visit sum of diameters of the non- missing lesions to the nadir sum of diameters excluding the lesions with missing measurements).

Example of scaling

Lesion	Longest diameter at nadir visit	Longest diameter at follow-up visit
1	7.2	7.1
2	6.7	6.4
3	4.3	4.0
4	8.6	8.5
5	2.5	Intervention
Sum	**29.3**	**26**

Lesion 5 has had an intervention at the follow-up visit.

The sum of lesions 1-4 at the follow-up is 26 mm. The sum of the corresponding lesions at baseline visit is 26.8 mm.

Scale up as follows to give an estimated TL sum of 28.4mm:

$$\frac{26}{26.8} \times 29.3 = 28.4\text{mm}$$

Lesions that split in two

If a TL splits in two, then the LDs of the split lesions should be summed and reported as the LD for the lesion that split.

Lesions that merge

If two TLs merge, then the LD of the merged lesion should be recorded for one of the TL sizes and the other TL size should be recorded as 0 mm.

Change in method of assessment of TLs

CT and MRI are the only methods of assessment that can be used within the trial. If a change in method of assessment occurs between CT and MRI, this will be considered acceptable and no adjustment within the programming is needed.

If a change in method involves clinical examination (e.g. CT changes to clinical examination or vice versa), any affected lesions should be treated as missing.

3.1.5 Site investigator assessment using RECIST 1.1: NTLs and new lesions

At each visit an overall assessment of the NTL response should be recorded by the investigator. This section provides the definitions of the criteria used to determine and record overall response for NTL at the investigational site at each visit.

NTL response will be derived based on the investigator's overall assessment of NTLs as follows:

Table 3 NTL Visit Responses

Visit Responses	Description
Complete Response (CR)	Disappearance of all NTLs present at baseline with all lymph nodes non-pathological in size (<10 mm short axis).
Progression (PD)	Unequivocal progression of existing NTLs. Unequivocal progression may be due to an important progression in one lesion only or in several lesions. In all cases the progression MUST be clinically significant for the physician to consider changing (or stopping) therapy.
Non-CR/Non-PD	Persistence of one or more NTLs with no evidence of progression.
Not Evaluable (NE)	Only relevant when one or some of the NTLs were not assessed and, in the investigator's opinion, they are not able to provide an evaluable overall NTL assessment at this visit. Note: For patients without TLs at baseline, this is relevant if any of the NTLs were not assessed at this visit and the progression criteria have not been met.
Not Applicable (NA)	Only relevant if there are no NTLs at baseline

To achieve 'unequivocal progression' on the basis of NTLs, there must be an overall level of substantial worsening in non-target disease such that, even in the presence of SD or PR in TLs, the overall tumour burden has increased sufficiently to merit a discontinuation of

therapy. A modest 'increase' in the size of one or more NTLs is usually not sufficient to qualify for unequivocal progression status.

Details of any new lesions will also be recorded with the date of assessment. The presence of one or more new lesions is assessed as progression.

A lesion identified at a follow up assessment in an anatomical location that was not scanned at baseline is considered a new lesion and will indicate disease progression.

The finding of a new lesion should be unequivocal: i.e. not attributable to differences in scanning technique, change in imaging modality or findings thought to represent something other than tumour.

New lesions will be identified via a Yes/No tick box. The absence and presence of new lesions at each visit should be listed alongside the TL and NTL visit responses.

A new lesion indicates progression so the overall visit response will be PD irrespective of the TL and NTL response.

If the question 'Any new lesions since baseline' has not been answered with Yes or No and the new lesion details are blank this is not evidence that no new lesions are present and should be treated as NE in the derivation of overall visit response.

Symptomatic deterioration is not a descriptor for progression of NTLs: it is a reason for stopping study therapy and will not be included in any assessment of NTLs.

Patients with 'symptomatic deterioration' requiring discontinuation of treatment without objective evidence of disease progression at that time should continue to undergo tumour assessments where possible until objective disease progression is observed.

3.1.6 Site investigator assessment using RECIST 1.1: overall visit response

Table 4 defines how the previously defined TL and NTL visit responses will be combined with new lesion information to give an overall visit response.

Table 4 Overall Visit Responses

Target Lesions	Non-target lesions	New Lesions	Overall Response
CR	CR	No	CR
CR	NA	No	CR
NA	CR	No	CR
CR	Non CR/Non PD	No	PR
CR	NE	No	PR
PR	Non PD or NE	No	PR

Table 4 Overall Visit Responses

Target Lesions	Non-target lesions	New Lesions	Overall Response
SD	Non PD or NE	No	SD
NA	Non CR/Non PD	No	SD (Non CR/Non PD)
NE	Non PD or NE	No	NE
NA	NE	No	NE
PD	Any	Yes or No	PD
Any	PD	Yes or No	PD
Any	Any	Yes	PD
NA	NA	No	NED

CR = complete response, PR = partial response, SD = stable disease, PD = progressive disease, NE = not evaluable, NED = no evidence of disease, NA = not applicable (only relevant if there were no TL/NTL at baseline).

3.2 Outcome Variables

All RECIST assessments, whether scheduled or unscheduled, will be included in the calculations. This is also regardless of whether a patient discontinues investigational product.

RECIST 1.1 outcomes (ie PFS, ORR etc.) will be derived using the overall visit responses and relevant dates from the BICR. This will be repeated using the programmatically derived overall visit responses from investigator RECIST 1.1 assessments.

3.2.1 Co-primary efficacy outcome endpoints

The co-primary endpoints are OS and PFS.

3.2.1.1 Overall survival (OS)

OS is defined as the time from the date of randomisation until death due to any cause (i.e., date of death or censoring – date of randomisation + 1). Any patient not known to have died at the time of analysis will be censored based on the last recorded date on which the patient was known to be alive (SUR_DAT, recorded within the SURVIVE module of the eCRF).

Note: Survival calls will be made in the week following the date of DCO for the analysis, and if patients are confirmed to be alive or if the death date is post the DCO date these patients will be censored at the date of DCO. The status of ongoing, withdrawn (from the study) and "lost to follow-up" patients at the time of the final OS analysis should be obtained by the site personnel by checking the patient's notes, hospital records, contacting the patient's general practitioner and checking publicly-available death registries. In the event that the patient has actively withdrawn consent to the processing of their personal data, the vital status of the patient can be obtained by site personnel from publicly-available resources where it is possible to do so under applicable local laws.

Note that for any OS analysis performed prior to the final OS analysis, in the absence of survival calls being made, it may be necessary to use all relevant CRF fields to determine the last recorded date on which the patient was known to be alive for those patients still on treatment (since the SURVIVE module is only completed for patients off treatment if a survival sweep is not performed). The last date for each individual subject is defined as the latest among the following dates recorded on the CRFs:

- AE start and stop dates
- Admission and discharge dates of hospitalization
- Study treatment date
- End of treatment date
- Laboratory test dates
- Date of vital signs
- Disease assessment dates on RECIST CRF
- Start and stop dates of alternative anticancer treatment
- Date last known alive on survival status CRF
- End of study date

3.2.1.2 Progression free survival (PFS)

PFS (per RECIST 1.1 as assessed by BICR) will be defined as the time from the date of randomisation until the date of the first objective disease progression or death (by any cause in the absence of progression) regardless of whether the patient withdraws from randomised therapy or receives another anti-cancer therapy prior to progression (i.e., date of PFS event or censoring – date of randomisation + 1). Patients who have not progressed or died at the time of analysis will be censored at the time of the latest date of assessment from their last RECIST 1.1 assessment. However, if the patient progresses or dies after 2 or more missed visits, the patient will be censored at the time of the latest RECIST 1.1 assessment prior to the two missed visits.

Given the scheduled visit assessment scheme (ie eight-weekly for the first 48 weeks then twelve-weekly thereafter) the definition of 2 missed visits will change. If the previous RECIST assessment is less than study day 274 (ie week 39) then two missing visits will equate to 18 weeks since the previous RECIST assessment, allowing for early and late visits (ie, 2 x 8 weeks + 1 week for an early assessment + 1 week for a late assessment = 18 weeks). If the two missed visits occur over the period when the scheduled frequency of RECIST assessments changes from eight-weekly to twelve-weekly this will equate to 22 weeks (ie,

take the average of 8 and 12 weeks which gives 10 weeks and then apply same rationale hence 2 x 10 weeks + 1 week for an early assessment + 1 week for a late assessment = 22 weeks). The time period for the previous RECIST assessment will be from study days 274 to 344 (ie week 39 to week 49). From week 49 onwards (when the scheduling changes to twelve-weekly assessments), two missing visits will equate to 26 weeks (ie 2 x 12 weeks + 1 week for an early assessment + 1 week for a late assessment = 26 weeks). If the patient has no visits or does not have baseline data they will be censored at study day 1 unless they die within 2 visits of baseline (16 weeks plus 1 week allowing for a late assessment within the visit window).

The PFS time will always be derived based on scan/assessment dates not visit dates.

PFS will also be obtained using the algorithm described above for the RECIST site investigator tumour data.

RECIST 1.1 assessments/scans contributing towards a particular visit may be performed on different dates. The following rules will be applied:

- The date of progression will be determined based on the earliest of the scan dates of the component that triggered the progression for the adjudicated reviewer selecting PD or of first reviewer where both select PD as a time point response and there is no adjudication for central review (BICR) data.

- For investigational assessments, the date of progression will be determined based on the earliest of the RECIST assessment/scan dates of the component that indicates progression

- When censoring a patient for PFS the patient will be censored at the latest of the dates contributing to a particular overall visit assessment.

Note: For TLs, only the latest scan date is recorded out of all scans performed at that assessment for the TLs and similarly for NTLs only the latest scan date is recorded out of all scans performed at that assessment for the NTLs. Additionally, PFS will be obtained using the algorithm described above for the RECIST BICR data, but following a modification whereby any objective disease progression must be confirmed by the next scheduled scan. The confirmatory scan must be no sooner than 4 weeks after the initial suspected progression. If disease progression is confirmed (or disease progression occurs and no further scans are recorded) then the date of progression will be when it was originally observed. Patients with a single disease progression and no further tumour assessment scans will be treated as PD in the analysis.

In the absence of clinically significant deterioration the investigational site is advised to continue the patient on study drug until progression has been confirmed.

For exploratory purposes, PFS will also be obtained using the irRECIST 1.1 data provided by BICR. Objective disease progressions also require confirmation under this approach.

Time to first subsequent therapy or death (TFST)

As a supportive summary to PFS, time to start of first subsequent therapy or death (TFST) will be derived. TFST is defined as the time from randomisation to the start date of the first subsequent therapy after discontinuation of treatment, or death, whichever is earlier (i.e. date of first subsequent anti-cancer therapy/death or censoring – date of randomisation + 1). Any patient not known to have had a first subsequent therapy will be censored at the last date that the patient was known not to have received a first subsequent therapy (obtained from the TTSCAPRX form). If a patient terminated the study for reason other than death before first subsequent therapy, these patients will be censored at the earliest of their known to be alive and termination dates. Patients not receiving randomised treatment would have TFST calculated in the same way, i.e. time from date of randomisation to the subsequent therapy or death. Note that the censoring for those patients without a subsequent therapy will align with their OS.

3.2.2 Secondary efficacy outcome endpoints

3.2.2.1 Proportion of patients alive at 24 months

The proportion of patients alive at 24 months (i.e., OS24) will be defined as the Kaplan-Meier estimate of OS at 24 months. In addition the proportion of patients alive at 12 months (i.e., OS12) will be presented. This will be defined as the Kaplan-Meier estimate of OS at 12 months.

3.2.2.2 Objective response rate

ORR (per RECIST 1.1 as assessed by the BICR) is defined as the number (%) of patients with at least 1 visit response of CR or PR and will be based on a subset of all randomised patients who have measurable disease at baseline per BICR. If the selected reviewer did not consider that a patient has measurable disease then this patient should not contribute to the denominator.

Data obtained up until progression, or the last assessment in the absence of progression, will be included in the assessment of ORR. Patients who go off treatment without progression, receive a subsequent therapy and then respond will not be included as responders in the ORR.

ORR will also be obtained using the algorithm described above for the RECIST site investigator data. The denominator for ORR will be all randomised patients with measurable disease at baseline per the site investigator (ie, the ITT population).

Additionally, ORR will be obtained using the algorithm described above from the RECIST BICR data, but following a modification where any objective progression requires confirmation. Therefore, data obtained up until confirmed progression (ie first progression date that is subsequently confirmed), or the last assessment in the absence of a confirmed progression, will be included in the assessment of ORR. Note that the response may be after an unconfirmed progression.

For exploratory purposes, ORR will also be obtained for the irRECIST 1.1 data provided by BICR.

Best objective response

Best objective response (BoR) is calculated based on the overall visit response from each RECIST assessment. It is the best response a patient has had following randomisation but prior to starting any subsequent cancer therapy and prior to RECIST progression or the last assessment in the absence of RECIST progression and subsequent cancer therapy.

Categorisation of BoR will be based on RECIST using the following response categories: CR, PR, SD, PD and NE.

BoR will be determined programmatically based on RECIST from the overall visit response at each visit using BICR data including all data up until the first progression, the start of any subsequent cancer therapy or the last assessment in the absence of progression.

For determination of a best response of SD, the earliest of the dates contributing towards a particular overall visit assessment will be used. SD should be recorded at least 8 weeks +/- 1 week, i.e. at least 49 days (to allow for the assessment window), after randomisation (i.e. study day 50). For CR/PR, the initial overall visit assessment which showed a response will use the latest of the dates contributing towards a particular overall visit assessment.

The denominator will be consistent with that used in the ORR analysis.

For patients whose PFS event is death, BoR will be calculated based upon all RECIST assessments prior to death.

For patients who die with no RECIST assessments, if the death occurs ≤17 weeks (i.e., 16 weeks + 1 week to allow for a late assessment within the assessment window) after randomisation, then BoR will be assigned to the progression (PD) category. For patients who die with no RECIST assessments, if the death occurs >17 weeks (i.e., 16 weeks + 1 week) after the date of first dose then BoR will be assigned to the NE category.

Progression events that have been censored due to them being more than two missed visits after the last assessment will not contribute to the BoR derivation.

Change in tumour size

For supportive purposes percentage change from baseline in tumour size will be derived at each scheduled tumour assessment visit (i.e., week 8, week 16 etc hereafter referred to as week X for convenience) using BICR data. Best percentage change from baseline in tumour size will also be derived as the biggest decrease or the smallest increase in tumour size from baseline.

This is based on RECIST TL measurements taken at baseline and at the timepoint of interest. Tumour size is defined as the sum of the longest diameters of the TLs based upon RECIST assessments. TLs are measurable tumour lesions. Baseline for RECIST is defined to be the last evaluable assessment prior to randomization. The change in TL tumour size at week X

will be obtained for each patient by taking the difference between the sum of the TLs at week X and the sum of the TLs at baseline. To obtain the percentage change in TL tumour size at week X the change in TL tumour size is divided by the sum of the TLs at baseline and multiplied by 100 (i.e. (week X - baseline) / baseline * 100).

Apply a window around the week X visit: Whenever tumour size data for the week X visit (Note: or visit at which progression was documented if before week X) is available then this should be used in the analysis. A windowing rule will be applied and will follow the protocol allowed visit window; therefore any RECIST scan performed within ± 1 week of the protocol scheduled visit will be used for that visit.

The above derivations will be programmed for the BICR based upon RECIST 1.1 assessments.

Measurements from the reviewer selected by the adjudicator will be used when adjudication for overall visit response has occurred, but in the case where no adjudication was required the measurements from the reviewer who reviewed the baseline scan first will be used for this analysis.

3.2.2.3 Duration of response

DoR (per RECIST 1.1 as assessed by the BICR) will be defined as the time from the date of first documented response until the first date of documented progression or death in the absence of disease progression, whichever is earlier (i.e., date of PFS event or censoring – date of first response + 1). The end of response should coincide with the date of progression or death from any cause used for the RECIST 1.1 PFS endpoint (Section 3.2.1.2). The denominator for DoR related analysis will be defined as described for ORR (see Section 3.2.2.2).

The time of the initial response will be defined as the latest of the dates contributing towards the first visit response of CR or PR.

If a patient does not progress following a response, then the corresponding DoR will be censored at the PFS censoring time.

DoR will not be defined for those patients who do not have documented response.

3.2.2.4 Proportion of patients alive and progression free at 12 months

The proportion of patients alive and progression free at 12 months (i.e., APF12) will be defined as the Kaplan-Meier estimate of PFS at 12 months.

3.2.2.5 Proportion of patients alive and progression free at 18 months

The proportion of patients alive and progression free at 18 months (i.e., APF18) will be defined as the Kaplan-Meier estimate of PFS at 18 months.

3.2.2.6 Time from randomisation to second progression or death (PFS2)

PFS2 will be defined as the time from the date of randomisation to the earlier of the progression event subsequent to that used for the PFS endpoint or death (i.e. date of progression event subsequent to that used for the PFS endpoint or date of death or censoring date whichever is earliest – date of randomisation + 1). The date of the first progression will be programmatically determined from investigator assessed data (See Section 3.1 for details.) The date of second progression will be recorded by the investigator and defined according to local standard clinical practice and may involve any of: objective radiological, symptomatic progression or death. RECIST assessments will not be collected for assessment of PFS2. The date of the PFS2 assessment and investigator opinion of progression status (progressed or non-progressed) at each assessment will be recorded in the electronic case report form (eCRF). Second progression status will be reviewed every 12 weeks following the progression used for the co-primary variable PFS (the first progression) and status recorded. Patients alive and for whom a second disease progression has not been observed should be censored at the last time known to be alive and without a second disease progression, i.e., censored at the latest of the PFS or PFS2 assessment date if the patient has not had a second progression or death.

Time to second subsequent therapy or death (TSST)

As a supportive summary to PFS2, time to start of second subsequent therapy or death (TSST) will be derived. TSST is defined as the time from randomisation to the earlier of the second subsequent cancer therapy start date following study treatment discontinuation, or death (i.e. date of second subsequent cancer therapy/death or censoring – date of randomisation + 1). Any patient not known to have had a second subsequent therapy will be censored at the last date that the patient was known not to have received a second subsequent anti-cancer therapy (obtained from the TTSCAPRX form). If a patient terminated the study for reason other than death before second subsequent therapy, these patients will be censored at the earliest of their last known to be alive and termination dates. Note that the censoring for those patients without a second subsequent therapy will align with their OS.

3.2.2.7 Time to death or distant metastasis

TTDM will be defined as the time from the date of randomisation until the first date of distant metastasis or death in the absence of distant metastasis. Distant metastasis is defined as any new lesion that is outside of the radiation field according to RECIST 1.1 or proven by biopsy. This definition is further refined as patients who enrol in the study have baseline scans performed only after completion of definitive chemoradiation. All known disease must have received a therapeutic dose of radiation prior to protocol entry. For this reason the TTDM endpoint will be determined from recurrent disease that occurs outside of the structures contained within the thorax, including lymph nodes, pulmonary, pleural, and mediastinal metastatic sites and excluding the heart. For the locations of distant metastases, see the Appendix. Patients who have not developed distant metastasis or died at the time of analysis will be censored at the time of the latest date of assessment from their last RECIST 1.1 assessment. However, if the patient has distant metastasis or dies after 2 or more missed visits, the patient will be censored at the time of the latest RECIST 1.1 assessment prior to the

2 missed visits. If the patient has no visits or does not have baseline data they will be censored at Day 1 unless they die within 2 visits of baseline.

3.3 Secondary Patient Reported Outcome Variables

PRO questionnaires will be assessed using the EORTC-QLQ-C30 with the EORTC-QLQ-LC-13 module (HRQoL with lung cancer specific additional concerns) and EQ-5D-5L. All items/questionnaires will be scored according to published scoring guidelines or the developer's guidelines, if published guidelines are not available. All PRO analyses will be based on the Intent-to-Treat (ITT) population, unless stated otherwise.

3.3.1 EORTC-QLQ-C30

The EORTC-QLQ-C30 consists of 30 questions which can be combined to produce 5 functional scales (physical, role, cognitive, emotional, social), 3 symptom scales (fatigue, pain, nausea/vomiting), 5 individual symptom items (dyspnoea, insomnia, appetite loss, constipation, diarrhoea) and a global measure of health status. The EORTC-QLQ-C30 will be scored according to the EORTC scoring manual (Fayers et al 1999). An outcome variable consisting of a score from 0 to 100 will be derived for each of the symptom scales/items, the functional scales and the global health status scale in the EORTC-QLQ-C30 according to the EORTC-QLQ-C30 Scoring Manual. Higher scores on the global health status and functioning scales indicate better health status/function but higher scores on symptom scales/items represent greater symptom severity.

Baseline will be defined as the last non-missing assessment prior to randomisation for symptoms and summaries.

The global health status/HRQoL will be assessed using the EORTC-QLQ-C30 global QoL scale which includes 2 items from the QLQ-C30: "How would you rate your overall health during the past week? (Item 29) and "How would you rate your overall QoL during the past week? (Item 30).

Definition of clinically meaningful changes

Changes in score compared with baseline will be evaluated. A minimum clinically meaningful change is defined as a change in the score from baseline of ≥10 for the EORTC-QLQ-C30 (Osoba et al 1998). For example, a clinically meaningful improvement in physical function (as assessed by EORTC-QLQ-C30) is defined as an increase in the score from baseline of ≥10, whereas a clinically meaningful deterioration is defined as a decrease in the score from baseline of ≥10. At each post-baseline assessment, the change in symptoms/functioning from baseline will be categorised as improvement, no change or deterioration as shown in Table 5.

Table 5 Visit response for EORTC-QLQ-C30

Score	Change from baseline	Visit response
Global health status/HRQoL	≥+10	Improvement
	≤-10	Deterioration
	Otherwise	No change
Symptom scales/items	≥+10	Deterioration
	≤-10	Improvement
	Otherwise	No change
Functional scales	≥+10	Improvement
	≤-10	Deterioration
	Otherwise	No change

EORTC-QLQ-C30 European Organization for Research and Treatment of Cancer 30-item core quality of life questionnaire.

For each subscale, if <50% of the subscale items are missing, then the subscale score will be divided by the number of non-missing items and multiplied by the total number of items on the subscales (Fayers et al 1999). If at least 50% of the items are missing, then that subscale will be treated as missing. Missing single items are treated as missing. The reason for any missing questionnaire will be identified and recorded. If there is evidence that the missing data are systematic, missing values will be handled to ensure that any possible bias is minimised.

For the visit level summaries of Improvement/Deterioration/No change then all patients with a baseline and post-baseline score will be included thus the denominator may differ from the time to deterioration and improvement rate endpoints derived below.

3.3.1.1 Time to symptom deterioration

For each of the symptoms scales/items in the EORTC-QLQ-C30, time to symptom deterioration will be defined as the time from randomisation until the date of the first clinically meaningful symptom deterioration (an increase in the score from baseline of ≥10) or death (by any cause) in the absence of a clinically meaningful symptom deterioration, regardless of whether the patient withdraws from study treatment or receives another anticancer therapy prior to symptom deterioration (i.e. date of symptom deterioration event or censoring whichever is earlier – date of randomisation + 1). Death will be included as an event only if the death occurs within 2 visits of the last PRO assessment where the symptom change could be evaluated.

Patients whose symptoms (as measured by EORTC-QLQ-C30) have not shown a clinically meaningful deterioration and who are alive at the time of the analysis will be censored at the time of their last PRO assessment where the symptom could be evaluated. Also, if symptoms deteriorate after 2 or more missed PRO assessment visits or the patient dies after 2 or more

missed PRO assessment visits, the patient will be censored at the time of the last PRO assessment where the symptom could be evaluated (prior to the two missed assessment visits). Given the scheduled visit assessment scheme (ie eight-weekly for the first 48 weeks then twelve-weekly thereafter) the definition of 2 missed visits will change. If the previous PRO assessment is less than study day 274 (ie week 39) then two missing visits will equate to 18 weeks since the previous PRO assessment, allowing for early and late visits. If the two missed visits occur over the period when the scheduled frequency of PRO assessments changes from eight-weekly to twelve-weekly this will equate to 22 weeks. The time period for the previous PRO assessment will be from study days 274 to 344 (ie week 39 to week 49). From week 49 onwards, two missing visits will equate to 26 weeks. If the patient has no evaluable visits or does not have baseline data they will be censored at day 1 unless they die within 2 visits of baseline (16 weeks plus 1 week allowing for a late assessment within the visit window).

The population for analysis of time to symptom deterioration will include a subset of the ITT population who have baseline scores ≤ 90.

3.3.1.2 Time to HR QoL/Function deterioration

For HR QoL/function, time to deterioration will be defined as the time from the date of randomisation until the date of the first clinically meaningful deterioration (a decrease in the function scales or the global health status/HR QoL from baseline of ≥10) or death (by any cause) in the absence of a clinically meaningful deterioration, regardless of whether the patient withdraws from study treatment or receives another anticancer therapy prior to HR QoL/function deterioration (i.e. date of HRQoL/function deterioration event or censoring whichever is earlier – date of randomisation + 1). Death will be included as an event only if the death occurs within 2 visits of the last PRO assessment where the HR QoL/function change could be evaluated.

Patients whose HR QoL/function (as measured by EORTC-QLQ-C30) have not shown a clinically meaningful deterioration and who are alive at the time of the analysis will be censored at the time of their last PRO assessment where the HR QoL/function could be evaluated. Also, if HR QoL/function deteriorates after 2 or more missed PRO assessment visits (using the same definitions for two missed visits as used in the 'Time to Symptom deterioration' derivation above) or the patient dies after 2 or more missed PRO assessment visits, the patient will be censored at the time of the last PRO assessment where the symptom could be evaluated. If the patient has no evaluable visits or does not have baseline data they will be censored at day 1 unless they die within 2 visits of baseline (16 weeks plus 1 week allowing for a late assessment within the visit window).

The population for analysis of time to HR QoL/function deterioration will include a subset of the ITT population who have baseline scores ≥ 10.

In the analysis, RECIST 1.1 progression will not be considered as HRQoL/function deterioration and data will not be affected by RECIST progression.

3.3.1.3 Symptom Improvement Rate

The symptom improvement rate will be defined as the number (%) of patients with 2 consecutive assessments at least 14 days apart that show a clinically meaningful improvement (a decrease from baseline score ≥10 for EORTC-QLQ-C30 symptom scales/items) in that symptom from baseline. The denominator will consist of a subset of the ITT population who have a baseline symptom score ≥10.

3.3.1.4 HR QoL/Function Improvement Rate

The HR QoL/function improvement rate will be defined as the number (%) of patients with 2 consecutive assessments at least 14 days apart that show a clinically meaningful improvement (an increase from baseline score ≥10 for EORTC-QLQ-C30 functional scales and global health status/HR QoL) in that scale from baseline. The denominator will consist of a subset of the ITT population who have a baseline HR QoL/function score ≤90.

3.3.2 EORTC-QLQ-LC-13

The EORTC-QLQ-LC-13 is a lung cancer specific module from the EORTC comprising 13 questions to assess lung cancer symptoms (cough, haemoptysis, dyspnoea, chest pain, arm/shoulder pain, and other pain), treatment related side-effects (sore mouth, dysphagia, peripheral neuropathy and alopecia) and pain medication. The LC-13 incorporates symptom scales including:

- Dyspnoea (multi-item scale based on 3 questions: were you short of breath when you rested; walked; climbed stairs)
- Cough: 1 item (how much did you cough?)
- Haemoptysis: 1 item (did you cough up blood?)
- Pain: 3 individual items (Have you had pain in your chest; your arm or shoulder; other parts of your body?)

The dyspnoea scale is only used if all 3 items have been scored; otherwise the items are treated as single-item measures. The scoring approach for the EORTC-QLQ-LC-13 is identical in principle to that for the symptom scales/single items of the EORTC-QLQ-C30.

Definition of clinically meaningful changes

Changes in score compared with baseline will be evaluated. A minimum clinically meaningful change is defined as an absolute change in the score from baseline of ≥10 for scales/items from the LC-13 (Osoba et al 1998). For example, a clinically meaningful deterioration or worsening in chest pain (as assessed by EORTC-QLQ-LC-13) is defined as an increase in the score from baseline of ≥10, whereas a clinically meaningful improvement is defined as a decrease in the score from baseline of ≥10. At each post-baseline assessment, the change in symptoms from baseline will be categorised as improvement, no change or deterioration as shown in Table 6.

Table 6 Visit response for L13

Score	Change from baseline	Visit response
Symptom scales/items	≥+10	Deterioration
	≤-10	Improvement
	Otherwise	No change

LC13 Lung Cancer Module.

For the visit level summaries of Improvement/Deterioration/No change then all patients with a baseline and post-baseline score will be included thus the denominator may differ from the time to deterioration and improvement rate endpoints derived below.

3.3.2.1 Time to symptom deterioration

For each of the symptoms scales/items in EORTC-QLQ-LC-13, time to symptom deterioration will be defined as the time from the date of randomisation until the date of the first clinically meaningful symptom deterioration (an increase in the score from baseline of ≥10) or death (by any cause) in the absence of a clinically meaningful symptom deterioration, regardless of whether the patient withdraws from study treatment or receives another anticancer therapy prior to symptom deterioration (i.e. date of symptom deterioration event or censoring – date of randomisation + 1). Death will be included as an event only if the death occurs within 2 visits of the last PRO assessment where the symptom change could be evaluated.

Patients whose symptoms (as measured by EORTC-QLQ-LC-13) have not shown a clinically meaningful deterioration and who are alive at the time of the analysis will be censored at the time of their last PRO assessment where the symptom could be evaluated. Also, if symptoms progress after 2 or more missed PRO assessment visits (using the same definitions for two missed visits as used in the 'Time to symptom deterioration' derivation in Section 3.3.1) or the patient dies after 2 or more missed PRO assessment visits, the patient will be censored at the time of the last PRO assessment where the symptom could be evaluated. If the patient has no evaluable visits or does not have baseline data they will be censored at day 1 unless they die within 2 visits of baseline (16 weeks plus 1 week allowing for a late assessment within the visit window).

The population for analysis of time to symptom deterioration will include a subset of the ITT population who have baseline scores ≤ 90 as a worsening of 10 points is not possible with a baseline value of >90.

In the analysis, RECIST 1.1 progression will not be considered as symptom deterioration and data will not be affected by RECIST progression.

3.3.2.2 Symptom Improvement Rate

The symptom improvement rate will be defined as the number (%) of patients with 2 consecutive assessments at least 14 days apart that show a clinically meaningful improvement (a decrease from baseline score ≥10 for LC-13 symptom scales/items) in that symptom from baseline. The denominator will consist of a subset of the ITT population who have a baseline symptom score ≥10 as a lessening of 10 points is not possible at a baseline score of <10.

3.3.3 PRO Compliance Rates

Summary measures of overall compliance and compliance over time will be derived for the EORTC-QLQ-C30 and EORTC-QLQ-LC13 respectively. These will be based upon:

- Received questionnaire = a questionnaire that has been received and has a completion date and at least one individual item completed.

- Expected questionnaire = a questionnaire that is expected to be completed at a scheduled assessment time e.g. a questionnaire from a patient who has not withdrawn from the study at the scheduled assessment time but excluding patients in countries with no available translation. For patients that have progressed, the latest of progression and safety follow-up will be used to assess whether the patient is still under HR QoL follow-up at the specified assessment time. Date of study discontinuation will be mapped to the nearest visit date to define the number of expected forms.

- Evaluable questionnaire = a questionnaire with a completion date and at least one subscale that is non-missing.

- Overall PRO compliance rate is defined as: Total number of evaluable questionnaires across all time points, divided by total number of questionnaires expected to be received across all time points multiplied by 100.

- Overall patient compliance rate is defined for each randomised treatment group as: Total number of patients with an evaluable baseline and at least one evaluable follow-up questionnaire (as defined above), divided by the total number of patients expected to have completed at least a baseline questionnaire multiplied by100.

Compliance over time will be calculated separately for each visit, including baseline, as the number of patients with an evaluable questionnaire at the time point (as defined above), divided by number of patients still expected to complete questionnaires. Similarly the evaluability rate over time will be calculated separately for each visit, including baseline, as the number of evaluable questionnaires (per definition above), divided by the number of received questionnaires.

3.4 Safety

Safety and tolerability will be assessed in terms of adverse events (AEs) (including serious adverse events [SAEs]), deaths, laboratory data, vital signs, electrocardiograms (ECGs) and exposure. These will be collected for all patients.

Data from the initial treatment period (i.e., the initial 12 months of treatment) on MEDI4736 will be compared against placebo in the main presentations of safety data and safety data from the re-treatment period may also be summarised separately (see Section 4.1). 'On treatment' will be defined as assessments between date of the first dose and 90 days following last dose of the study treatment (MEDI4736/Placebo) unless otherwise specified. Note that for one version of the safety outputs the period of time after the administration of subsequent therapy will not be considered 'on treatment' (see further Section 0).

The Safety analysis set will be used for reporting of safety data.

3.4.1 Adverse events

AEs and SAEs will be collected throughout the study, from date of informed consent to 90 days after the last dose of MEDI4736 or placebo. The Medical Dictionary for Regulatory Activities (MedDRA) will be used to code the AEs. AEs will be graded according to the National Cancer Institute Common Terminology Criteria for AEs (CTCAE Version 4.03). A treatment emergent adverse event (TEAE) is an AE with an onset date or a pre-existing AE worsening following the first dose of study treatment through to 90 days after the last dose of MEDI4736 or placebo.

Adverse events that have missing causality (after data querying) will be assumed to be related to study drug.

Other significant adverse events

During the evaluation of the AE data, an AstraZeneca medically qualified expert will review the list of AEs that were not reported as SAEs and 'Discontinuation of Investigational Product due to Adverse Events' (DAEs). Based on the expert's judgment, significant adverse events of particular clinical importance may, after consultation with the Global Patient Safety Physician, be considered other significant adverse events (OAEs) and reported as such in the CSR. A similar review of laboratory/vital signs/ECG data will be performed for identification of OAEs.

Examples of these are marked haematological and other laboratory abnormalities, and certain events that lead to intervention (other than those already classified as serious) or significant additional treatment.

AEs of special interest

Some clinical concepts (including some selected individual preferred terms [PTs] and higher level terms [HLTs]) have been considered "AEs of special interest" (AESI) to the MEDI4736

program. AESIs represent pre-specified risks which are considered to be of importance to a clinical development program.

These AESIs identified as a list of categories provided by the clinical team. Other categories may be added as necessary or existing terms may be merged. An AstraZeneca medically qualified expert, after consultation with the Global Patient Safety Physician, has reviewed the AEs of interest and identified which preferred terms contribute to each AESI. A further review will take place prior to DBL to ensure any further terms not already included are captured within the categories.

3.4.2 Laboratory data

Laboratory data will be collected throughout the study, from screening to follow-up visit as described in the CSP. Blood and urine samples for determination of haematology, clinical chemistry, and urinalysis will be collected as described in Section 6.4.5 of the CSP. For the definition of baseline and the derivation of post baseline visit values considering visit window and how to handle multiple records, derivation rules as described in Section 4.1.3 below will be used.

Change from baseline in haematology and clinical chemistry variables will be calculated for each post-dose visit on treatment. CTCAE grades will be defined at each visit according to the CTCAE grade criteria using local or project ranges as required, after conversion of lab result to corresponding preferred units. The following parameters have CTCAE grades defined for both high and low values: Potassium, Sodium, Magnesium, Glucose and Corrected calcium so high and low CTCAE grades will be calculated.

Corrected calcium will be derived during creation of the reporting database using the following formula:

Corrected calcium (mmol/L) = Total Calcium (mmol/L) + ([40 – Albumin (G/L)] x 0.02)

Creatinine clearance will be derived according to the Cockroft-Gault formula (Cockcroft and Gault 1976).
Males:

$$\text{Creatinine CL} = \frac{\text{Weight (kg) x (140 - Age)}}{\text{72 x serum creatinine (mg/dL)}}$$

Females:

$$\text{Creatinine CL} = \frac{\text{Weight (kg) x (140 - Age)}}{\text{72 x serum creatinine (mg/dL)}} \text{ x } 0.85$$

Absolute values will be compared to the project reference range and classified as low (below range), normal (within range or on limits of range) and high (above range).

The maximum or minimum on-treatment value (depending on the direction of an adverse effect) will be defined for each laboratory parameter as the maximum (or minimum) post-baseline value at any time.

Local reference ranges will be used for the primary interpretation of laboratory data at the local laboratory. Project reference ranges will be used throughout for reporting purposes. The denominator used in laboratory summaries of CTCAE grades will only include evaluable patients, in other words those who had sufficient data to have the possibility of an abnormality.

For example:

- If a CTCAE criterion involves a change from baseline, evaluable patients would have both a baseline and at least 1 post-dose value recorded.
- If a CTCAE criterion does not consider changes from baseline, to be evaluable the patient need only have 1 post-baseline value recorded.

3.4.3 ECGs

ECG data obtained up until the 30 days from date of last dose of study treatment will be used for reporting. For derivation of post baseline visit values considering visit window and to handle multiple records present in any visit window, derivation rules as described in Section 4.1.3 below will be used. The denominator in ECGs data should include only those patients with recorded data.

At each time point the Investigator's assessment of the ECG will be collected locally. Heart rate, duration of QRS complex, RR, PR and QT intervals will be collected centrally via a digital read. This digital copy of all ECGs will be held centrally by a central ECG provider, and the data from this review will be stored for analysis if necessary at the end of the study. If it is necessary to analyse this data then QTcF (Fridericia) will be calculated programmatically using the reported ECG values (RR and QT).

$$\text{QTcF (msec)} = \frac{\text{QT (msec)}}{\sqrt[3]{\text{RR (sec)}}}$$

For triplicate ECGs, the mean of the three ECG assessments will be used to determine the value at that time point.

3.4.4 Vital signs

Vital signs data obtained up until the 30 days from date of last dose of study treatment will be used for reporting. Change from baseline in vital signs variables will be calculated for each

post-dose visit on treatment. For derivation of post baseline visit values considering visit window and to handle multiple records, derivation rules as described in Section 4.1.3 below will be used. The denominator in vital signs data should include only those patients with recorded data.

3.4.5 Time to first subsequent therapy from discontinuation of study treatment

Time to subsequent therapy from date of last dose is defined as the time from the date of discontinuation of study treatment to the start date of the first subsequent therapy after treatment discontinuation. Any patient not known to have had a first subsequent therapy will not have this calculation performed.

3.4.6 Treatment exposure

Exposure will be defined as follows for the initial treatment period and for the re-treatment period:

Total (or intended) exposure of study medication,

- Total (or intended) exposure = min(last dose date where dose>0mg+13 days, date of death, date of DCO) – first dose date +1

Actual exposure of MEDI4736/placebo

- Actual exposure = intended exposure – total duration of dose delays, where intended exposure will be calculated as above.

Dose reductions are not permitted per the CSP. The actual exposure calculation makes no adjustment for any dose reductions that may have occurred.

Calculation of duration of dose delays (for actual exposure):

- Since patients will receive 10 mg/kg via IV infusion q2w for up to 12 months (up to 26 doses), the duration of dose delays will be calculated as follows:

  For all dosing dates:

  Total duration of dose delays= Sum of (Date of the dose - Date of previous dose – 14 days)

  Thus, if no delays were encountered, the duration would sum up to 0, since infusions were done every two weeks.

3.4.7 Dose intensity

Dose intensity will be derived separately for the initial treatment period and the re-treatment period. Relative dose intensity (RDI) is the percentage of the actual dose intensity delivered relative to the intended dose intensity through to treatment discontinuation.

RDI will be defined as follows:

- RDI = 100% * d/D, where d is the actual cumulative dose delivered up the actual last day of dosing and D is the intended cumulative dose up to the actual last day of dosing. D is the total dose that would be delivered, if there were no modification to dose or schedule. When accounting for the calculation of intended cumulative dose 3 days should be added to the date of last dose to reflect the protocol allowed window for dosing for all treatments

When deriving actual dose administered the volume before and after infusion will also be considered.

For the Study medication treatment, where the last dose is on the week 50 visit and if there are scans post week 50, the censoring of progression should occur at week 50 for the purposes of RDI.

Example of dose intensity for MEDI4736

Table 7 Dose intensity scenarios

		Study Day									
RDI	Patient	1	15	29	43	57	71	85	99	113	
100%	1	X	X	X	X	X	X	X	X	X	PD
100%	2	X	X	X	X	X	X	X	X[D]		PD
56%	3	X	X			X	O	X		X	PD
67%	4	X	X	O	X	X	X	O	X	O	PD

X: Dose of 10mg/kg taken; O: Dose missed; [D]: Dose discontinued; PD: Progressive Disease

Patients 1-4 progressed on Day 115, so the intended dose through to progression was 9 * 10mg/kg of MEDI4736 = 90mg/kg.

Patient 1 received a total of 90 mg/kg of MEDI4736, whereas other patients received less treatment due to:

- Early stopping prior to PD (Patient 2)

- Dosing delays (Patient 3)
- Missed doses (Patient 4)

The examples of Patients 2 and 4 illustrate that for RDI, the end of actual dosing period is calculated based on the smallest recovery period after the last non-zero dose.

Patient 1: RDI = (9 * 10 mg/kg)/ 90 mg/kg = 100%

Patient 2: RDI = (8 * 10 mg/kg)/ 80 mg/kg = 100%

Patient 3: RDI = (5 * 10 mg/kg)/ 90 mg/kg = 56%

Patient 4: RDI = (6 * 10 mg/kg)/ 90 mg/m^2 = 67%

3.5 Pharmacokinetic and Immunogenicity Variables

Analyses to evaluate the pharmacokinetics and immunogenicity of MEDI4736 will be performed by AstraZeneca/MedImmune Clinical Pharmacology group or designee.

3.5.1 PK non-compartmental analysis

The actual sampling times will be used in the PK calculations. Pharmacokinetic concentration data and summary statistics will be tabulated. Individual and mean blood MEDI4736 concentration-time profiles will be generated. PK parameters will be determined using standard non-compartmental methods. The PK parameters will be determined after the first and steady state doses: peak and trough concentration (as data allow). Samples below the lower limit of quantification will be treated as missing in the analyses.

3.5.2 Population PK and exposure-response/safety analysis

A population PK model will be developed using a non-linear mixed-effects modelling approach in patients with NSCLC where possible. The impact of physiologically-relevant patient characteristics (covariates) and disease on PK will be evaluated. The relationship between PK exposure and the effect on safety and efficacy will be evaluated, if appropriate. The results of such an analysis will be reported in a separate report.

3.5.3 Immunogenicity analysis

Immunogenicity results will be analysed descriptively by summarizing the number and percentage of patients who develop detectable anti-MEDI4736 antibodies. The immunogenicity titre will be reported for samples confirmed positive for the presence of anti-MEDI4736 antibodies. Summaries will be based upon all patients from the safety population. The effect of immunogenicity on PK, PDx, efficacy and safety will be evaluated, but such analyses, if applicable, will be reported in a separate report.

3.6 Exploratory Variables

The following exploratory variables will be included in the CSR.

3.6.1 EQ-5D-5L

As one of the exploratory objectives, the EQ-5D-5L health state utility index will be used to explore the impact of treatment and disease state on health state utility.

The EQ-5D is a standardised measure of health status developed by the EuroQol Group in order to provide a simple, generic measure of health for clinical and economic appraisal. Applicable to a wide range of health conditions and treatments, it provides a simple descriptive profile and a single index value for health status that can be used in the clinical and economic evaluation of health care.

The EQ-5D-5L index comprises 5 dimensions of health (mobility, self-care, usual activities, pain/discomfort and anxiety/depression). For each dimension, respondents select which statement best describes their health on that day from a possible 5 options of increasing levels of severity (no problems, slight problems, moderate problems, severe problems and unable to/extreme problems). A unique EQ-5D health state is referred to by a 5 digit code allowing for a total of 3125 health states. For example, state 11111 indicates no problems on any of the 5 dimensions. These data will be converted into a weighted health state index by applying scores from EQ-5D value sets elicited from general population samples (the base case will be the United Kingdom valuation set, with other country value sets applied in scenario analyses). Where values sets are not available, the EQ-5D-5L to EQ-5D-3L crosswalk (Oemar and Janseen 2013) will be applied. In addition to the descriptive system, respondents also assess their health on the day of assessment on a visual analogue scale, ranging from 0 (worst imaginable health) to 100 (best imaginable health). This score is reported separately.

The evaluable population will comprise a subset of the ITT analysis set who have a baseline EQ-5D-5L assessment.

Compliance rate of EQ-5D-5L will be analysed similarly using the method stated in Section 3.3.3.

3.6.2 Health Resource Use

Health resource use outcome variables include the following:

- Length of hospital stay
- Reasons for hospitalisation
- Length of any time spent in an intensive care unit (ICU)

The length of hospital stay will be calculated as the difference between the date of hospital discharge (or death date) and the start date of hospitalisation or start of study drug if the start

of study drug is after start date of hospitalisation (length of hospital stay = end date of hospitalisation – start date of hospitalisation or start date of study drug whichever is later + 1). For patients with missing discharge date, the last day with available data will be used as the end date of hospitalisation.

For patients who were hospitalised more than one time during the study period, the length of hospital stay will be the sum of all durations of hospital stays.

The length of ICU stay will be calculated using the same method as detailed above for the length of hospital stay.

4. ANALYSIS METHODS

There will be 1 treatment comparison of interest:

- MEDI4736 10 mg/kg vs placebo

The co-primary endpoints are OS and PFS using RECIST 1.1. The study has been sized to characterise the OS and PFS benefit of MEDI4736 relative to placebo.

Results of all statistical analysis will be presented using a 95% CI and 2-sided p-value, unless otherwise stated.

4.1 General Principles

Efficacy and HRQoL data will be summarised and analysed on the ITT analysis set. Safety and treatment exposure data will be summarised based upon the safety analysis set. Study population and demography data will be summarised based upon the ITT analysis set.

Efficacy from the re-treatment period may be summarised separately for the site investigator data. For the site investigator data, any derivations relative to baseline (e.g. day, RECIST derivations) in the re-treatment period will be relative to the baseline scan prior to the re-treatment. To avoid biasing the blinded central reviewers (i.e., when determining progression) the baseline scans prior to re-treatment will not be identified (see further the BICR charter) thus no efficacy data will be summarised from the re-treatment period based upon BICR.

Safety data will be summarised from the initial treatment period. Safety data from the re-treatment period may also be summarised via a small set of headline summaries should there be sufficient number of patients re-treated to warrant this. Any safety summaries representing the re-treatment period will be based upon a subset of the safety analysis set representing patients who have had at least one dose of study treatment in the re-treatment period.

The below mentioned general principles will be followed throughout the study:

- Descriptive statistics will be used for all variables, as appropriate. Continuous variables will be summarised by the number of observations, mean, standard deviation, median,

minimum, and maximum. For log transformed data it is more appropriate to present geometric mean, coefficient of variation (CV), median, minimum and maximum. Categorical variables will be summarised by frequency counts and percentages for each category.

- Unless otherwise stated, percentages will be calculated out of the population total and for each treatment group.

- For continuous data the mean and median will be rounded to 1 additional decimal place compared to the original data. The standard deviation will be rounded to 2 additional decimal places compared to the original data. Minimum and maximum will be displayed with the same accuracy as the original data.

- For categorical data, percentages will be rounded to 1 decimal place.

- SAS® version 9.2 will be used for all analyses.

4.1.1 Baseline

In general, for efficacy and PRO endpoints the last observed measurement prior to randomization will be considered the baseline measurement. However, if an evaluable assessment is only available after randomization but before the first dose of randomized treatment then this assessment will be used as baseline. For safety endpoints the last observation before the first dose of study treatment will be considered the baseline measurement unless otherwise specified. For assessments on the day of first dose where time is not captured, a nominal pre-dose indicator, if available, will serve as sufficient evidence that the assessment occurred prior to first dose.

Assessments on the day of the first dose where neither time nor a nominal pre-dose indicator are captured will be considered prior to the first dose if such procedures are required by the protocol to be conducted before the first dose.

In all summaries change from baseline variables will be calculated as the post-treatment value minus the value at baseline. The % change from baseline will be calculated as (post-baseline value - baseline value) / baseline value x 100.

4.1.2 Multiplicity

The multiple testing procedure will define which significance levels should be applied to the interpretation of the raw p-values for the 2 primary endpoints of PFS and OS and the key secondary endpoints of OS24 and ORR.

There will be 4 DCO time points in the study. The DCO for the interim PFS analysis (first analysis) will be done when 367 PFS events have occurred (52% maturity), approximately 30 months after the first patient is randomised. The DCO for the primary PFS analysis (second analysis) and interim OS analysis will be done when at least 458 PFS events have occurred

(65% maturity) at approximately 36 months after first patient is randomised. It is expected that approximately 285 death events (41% maturity) will be available for the interim OS analysis. The third analysis data cut-off will occur at the time of the second OS interim analysis when it is expected that 393 OS events have occurred (56% maturity, approximately 47 months after the first patient is randomised). The DCO for the primary OS analysis (fourth analysis) will be done when 491 OS events have occurred (70% maturity) at approximately 62 months.

The overall 5% type 1 error will be split between the co-primary endpoints OS and PFS. An alpha level of 2.5% will be allocated to OS analysis and an alpha level of 2.5% will be allocated to the PFS analysis. An interim PFS analysis for superiority will occur when approximately 367 PFS events have occurred and the primary PFS analysis will be performed when 458 PFS events have accumulated. The 2.5% alpha level allocated to PFS will be controlled at the interim and primary time point by using the Lan DeMets (Lan and DeMets 1983) spending function that approximates an O'Brien Fleming approach, where the significance level applied at the interim depends upon the proportion of information available. If 80% of the PFS events required at the time of the primary PFS analysis is available at the time of the interim (i.e., 367/458 PFS events have occurred), the 2-sided significance level to be applied for the PFS interim analysis would be 1.04% and the 2-sided significance level to be applied for the primary PFS analysis would be 2.19%.

An interim OS analysis for superiority will occur at the time of the primary PFS analysis. The 2.5% alpha level allocated to OS will be controlled at the 2 interim and primary time point by using the Lan DeMets (Lan and DeMets 1983) spending function that approximates an O'Brien Fleming approach, where the significance level applied at the interim depends upon the proportion of information available. For example, if 58% or 80% of OS events required at the time of the primary OS analysis is available at the time of the interim (i.e., 285/491 or 393/491 events have occurred), the 2-sided significance level to be applied for the OS interim analysis would be 0.21% and 0.98% and the 2-sided significance level to be applied for the primary OS analysis would be 2.17%.

The family wise error rate (FWER) will be strongly controlled for the testing of PFS, OS, OS24, and ORR including interim analyses by applying the testing procedure illustrated in the Figure 2 below.

The PFS and OS will be tested according to the group sequential Holm fixed (GSHf) procedure described in Ye et al 2012. In this procedure, the testing will start from PFS, if the testing for PFS is significant at the alpha level specified in Figure 2 at either interim or primary analysis the full 2.5% alpha level for PFS can be propagated to the testing of OS, which means that the OS will be tested at an overall alpha level of 5%. Following the GSHf procedure the alpha level for the interim analyses of OS will not be changed but the alpha level for the primary analysis of OS will be recalculated based on 5% alpha level overall with interim alpha levels as specified in Figure 2. For example, if the information fraction for the interim analyses is as specified (80% for PFS interim and 58%/80% for OS interims) the alpha level for the primary OS analysis will be 4.85% so that the FWER for OS is 5% accounting

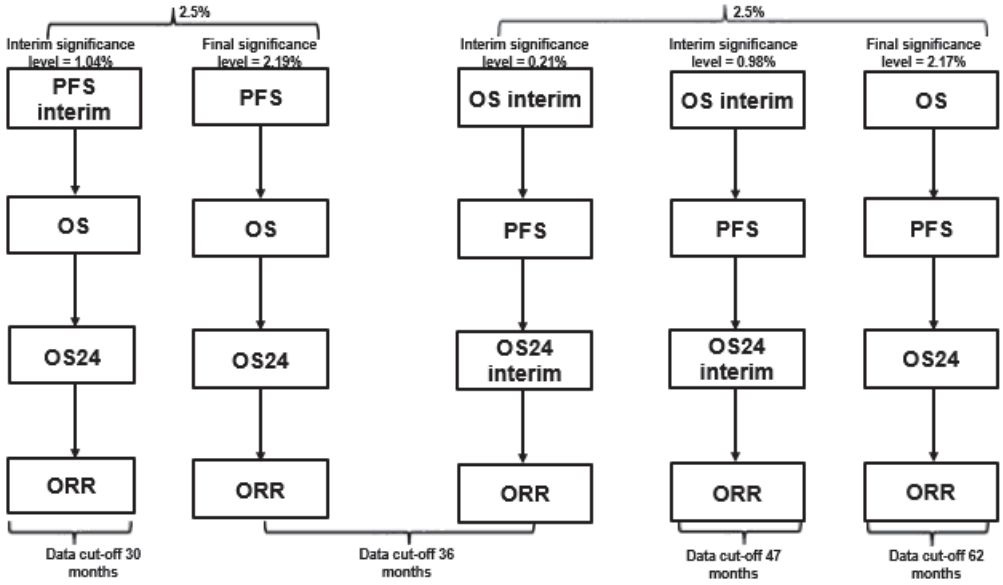
for interim analyses, given testing of PFS is significant at interim (alpha=1.04%) or at primary analysis (alpha=2.19%). The alpha level for the interim OS analyses will remain at 0.21% and 0.98%.

If the testing for PFS is not significant at either the interim (alpha=1.04%) or at the primary analysis (alpha=2.19%), the OS will be tested at the levels specified in Figure 2. In this case if the testing for OS is significant at any interim analysis or primary analysis the full 2.5% alpha level for OS can be propagated to the testing of PFS, which means that the PFS will be tested at an overall alpha level of 5%. Following GSHf, the alpha level for the interim analysis of PFS will not be changed but the alpha level for the primary analysis of PFS will be recalculated based on 5% alpha level overall with interim alpha levels as specified in Figure 2. For example, if the information fraction for the interim analyses is as specified (80% for PFS interim and 58%/80% for OS interim) the alpha level for the primary PFS analysis will be 4.88%, given testing of OS is significant at interim (alpha=0.21% and 0.98%) or at primary analysis (alpha=2.17%). The alpha level for the interim PFS analysis will remain at 1.04%.

OS24 and ORR will not be tested unless the null hypotheses for both PFS and OS are rejected according to the GSHf procedure described above. After both PFS and OS are rejected the OS24 will be tested with 5% alpha level controlled at the interim and primary time point. The interim analyses of OS24 will use the same interim alpha level for OS at the same analysis point. The alpha for primary analysis of OS24 will be calculated based on interim alpha levels and information fractions at the interim so that the alpha level for OS24 is controlled at 5% overall. ORR will only be tested after the null hypothesis of PFS, OS, and OS24 are all rejected and ORR will be tested at a 5% level.

The FWER is strongly controlled at 5% for PFS and OS according to Ye et al 2012. The FWER for PFS, OS, and OS24 is also strongly controlled at 5% due to the fix-sequence testing nature and according to Glimm et al 2009. Finally the FWER is still strongly controlled after adding the fix-sequence testing of ORR. So the FWER is strongly controlled at 5% for the testing of all four endpoints. Figure 2 shows the multiple testing framework.

Figure 2 Multiple testing procedures for controlling the type 1 error rate



4.1.3 Visit window for safety and PRO assessments

Time windows will need defining for any presentations that summarise values by visit. The following conventions should also apply:

- The time windows should be exhaustive so that data recorded at any time point has the potential to be summarised. Inclusion within the time window should be based on the actual date and not the intended date of the visit.

- All unscheduled visit data should have the potential to be included in the summaries.

- The window for the visits following baseline will be constructed in such a way that the upper limit of the interval falls half way between the two visits (the lower limit of the first post-baseline visit will be Day 2). If an even number of days exists between two consecutive visits then the upper limit will be taken as the midpoint value minus 1 day.

  For example, the visit windows for vital signs data are:

  – Day 15, visit window 2 - 21

  – Day 29, visit window 22 – 35

  – Day 43, visit window 36 - 49

  – Day 57, visit window 50 – 63

  - Day 71, visit window 64 – 77
  - Day 85, visit window 78 – 91

- For summaries showing the maximum or minimum values, the maximum/minimum value recorded on treatment will be used (regardless of where it falls in an interval).
- Listings should display all values contributing to a time point for a patient.
- For visit based summaries:
  - If there is more than one value per patient within a time window then the closest value to the scheduled visit date should be used, or the earlier in the event the values are equidistant from the nominal visit date. If there are two values recorded on the same day and the parameter is CTCAE gradeable then the record with the highest toxicity grade should be used. Alternatively, if there are two records recorded on the same day and the toxicity grade is the same (or is not calculated for the parameter) then the average of the two records should be used. The listings should highlight the value for that patient that went into the summary table, wherever feasible. Note: in summaries of extreme values all on-treatment values collected are used including those collected at unscheduled visits.
  - To prevent very large tables or plots being produced that contain many cells with meaningless data, for each treatment group visit data should only be summarised if the number of observations is greater than the minimum of 20 and > 1/3 of patients dosed.
- For summaries at a patient level, all values should be included, regardless of whether they appear in a corresponding visit based summary, when deriving a patient level statistic such as a maximum.
- Baseline will be defined as the last non-missing measurement prior to the first dose of study treatment. For the re-treatment period baseline is similarly defined as the last non-missing measurement prior to the first dose on the re-treatment period. For laboratory data, any assessments made on day 1 will be considered pre-dose. If there are two visits equally eligible to assess patient status at baseline (e.g., screening and baseline assessments both on the same date prior to first dose with no washout or other intervention in the screening period) with assessment time missing, the average can be taken as a baseline value. For non-numeric laboratory tests (ie some of the urinalysis parameters) where taking an average is not possible then the best value would be taken as baseline as this is the most conservative. In the scenario where there are two assessments on day 1, one with time recorded and the other without time recorded, the one with time recorded would be selected as baseline.

- Where safety data are summarised over time, study day will be calculated in relation to date of first treatment.

Missing safety data will generally not be imputed. However, safety assessment values of the form of "< x" (i.e., below the lower limit of quantification) or > x (i.e., above the upper limit of quantification) will be imputed as "x" in the calculation of summary statistics but displayed as "< x" or "> x" in the listings. Additionally, adverse events that have missing causality (after data querying) will be assumed to be related to study drug.

4.2 Analysis Methods

Table 8 details which endpoints are to be subject to formal statistical analysis, together with pre-planned sensitivity analyses making clear which analysis is regarded as primary for that endpoint.

Table 8 Formal statistical analyses to be conducted and pre-planned sensitivity analyses

Endpoints Analysed	Notes
Overall Survival	• Co-primary analysis using a stratified log-rank test • Sensitivity analysis using a Kaplan-Meier plot of time to censoring where the censoring indicator of the primary analysis is reversed – attrition bias • Additional analysis using Cox proportional hazards models to determine the effect of covariates on the HR estimate • Additional analysis using Cox proportional hazards models to determine the consistency of treatment effect between subgroups via the approach of Gail and Simon 1985. • Subgroup analysis using Cox proportional hazards model

Table 8 Formal statistical analyses to be conducted and pre-planned sensitivity analyses

Endpoints Analysed	Notes

Progression Free Survival	• Co-primary analysis using stratified log-rank test using BICR data (RECIST 1.1)
	• Sensitivity analyses using BICR data (RECIST 1.1)
	  − Interval censored analysis – evaluation time bias
	  − Analysis using alternative censoring rules – attrition bias
	• Sensitivity analysis stratified log-rank test using site investigator tumour data (RECIST 1.1) – ascertainment bias
	• Sensitivity analysis stratified log-rank test using BICR data (RECIST 1.1, modified for confirmation of progression) – confirmation bias
	• Additional analysis using Cox proportional hazards models to determine the effect of covariates on the HR estimate
	• Additional analysis using Cox proportional hazards models to determine the consistency of treatment effect between subgroups via the approach of Gail and Simon 1985.
	• Subgroup analysis using Cox proportional hazards model
	• Exploratory analysis stratified log-rank test using BICR tumour data (irRECIST)

Proportion of patients alive at 24 months	Kaplan-Meier estimates of survival at 24 months and p-value (following the method described by Klein et al 2007)
Objective Response Rate	Fisher's exact test using BICR data (RECIST 1.1) Sensitivity analysis using the Fisher's exact test using site investigator tumour data (RECIST 1.1) Sensitivity analysis using the Fisher's exact test using BICR tumour data (RECIST 1.1, modified for confirmation of progression) Exploratory analysis using the Fisher's exact test using BICR tumour data (irRECIST)
Duration of Response	Analysis following the method described by Ellis et al 2008 using BICR tumour data (RECIST 1.1)
Proportion of patients alive and progression free at 12 and18 months	Kaplan Meier estimates of progression free survival at 12 and 18 months
Time from randomisation to second progression	Stratified log-rank test
Time to death or distant metastasis	Stratified log-rank test using site BICR tumour data (RECIST 1.1)

Table 8 Formal statistical analyses to be conducted and pre-planned sensitivity analyses

Endpoints Analysed	Notes
Symptom improvement rate (EORTC QLQ-C30 and LC13 endpoints)	Logistic regression
HR QoL/Function improvement rate (EORTC QLQ-C30 endpoints)	Logistic regression
Time to HR QoL/Function deterioration (EORTC QLQ-C30 endpoints)	Stratified log-rank test
Time to symptom deterioration (EORTC QLQ-C30 and LC13 endpoints)	Stratified log-rank test
Change from baseline (EORTC QLQ C30 and LC13; EQ-5D-5L health state utility values and Visual Analogue Scale)	Mixed model repeated measures analysis

BICR Blinded Independent Central Review; EORTC QLQ-C30 European Organisation for Research and Treatment of Cancer 30-item core quality of life questionnaire; EQ-5D-5L EuroQoL 5 dimension, 5 level health state utility index; irRC Immune-related response criteria; LC13 Lung Cancer Module; HR QoL Quality of Life; RECIST Response Evaluation Criteria In Solid Tumours.

All outputs will be summarised by treatment arm as randomised for all ITT patients.

4.2.1 Co-primary efficacy endpoints

4.2.1.1 Overall survival

The primary analysis of the co-primary endpoint, OS, will occur when approximately 491 deaths have occurred (approximately 70% maturity). OS will be analysed using a stratified log-rank test adjusting for age at randomisation (<65 vs ≥65 years of age), sex (male vs female), and smoking history (smoker vs non-smoker) for generation of the p-value and using the Breslow approach for handling ties (Breslow, 1974).

The effect of treatment will be estimated by the hazard ratio (HR) together with its corresponding (1-alpha adjusted)% CI and p-value for the ITT population.

The covariates in the statistical modelling will be based on the values entered into IVRS at randomisation, even if it is subsequently discovered that these values were incorrect. In addition a sensitivity analysis will be produced which use the covariates recorded in the CRF.

The HR and CI can be estimated from the stratified log-rank as follows (Berry et al 1991, Collett 2003, Sellke and Siegmund 1983):

$$HR = \exp\left(\frac{U}{V}\right)$$

$$95\%\ \text{CI for HR} = \left(\exp\left\{\frac{U}{V} - \frac{1.96}{\sqrt{V}}\right\}, \exp\left\{\frac{U}{V} + \frac{1.96}{\sqrt{V}}\right\}\right)$$

Where $U = \sum_k U_k = \sum_k \sum_i (d_{1ki}, -e_{1ki})$ is the stratified log-rank test statistic obtained from the SAS LIFTEST procedure, $\sqrt{V} = \sqrt{\sum_k V_k}$, is its standard deviation, k denotes the stratum and d_{1ki} and e_{1ki} are the observed and expected events in Group 1, stratum k.

Kaplan-Meier plots of OS will be presented by treatment arm. Summaries of the number and percentage of patients who have died, those still in survival follow-up, those lost to follow-up and those who have withdrawn consent will be provided along with the median OS for each treatment. The assumption of proportionality will be assessed.

Proportional hazards will be tested firstly by examining plots of complementary log-log (event times) vs log (time) and, if these raise concerns, by fitting a time-dependent covariate to assess the extent to which this represents random variation. If a lack of proportionality is evident, the variation in treatment effect will be described by presenting piecewise HR calculated over distinct time-periods. In such circumstances, the HR can still be meaningfully interpreted as an average HR over time unless there is extensive crossing of the survival curves. If lack of proportionality is found, this may be is a result of treatment-by-covariate interactions, which will be investigated.

Sensitivity analyses

A sensitivity analysis for OS will examine the censoring patterns to rule out attrition bias, achieved by a Kaplan-Meier plot of time to censoring where the censoring indicator of OS is reversed.

The number of patients prematurely censored will be summarised by treatment arm. A patient would be defined as prematurely censored if their survival status was not defined at the DCO.

In addition, duration of follow-up will be summarised using medians:

- In censored patients who are alive at DCO only: Time from randomisation to date of censoring (date last known to be alive) for each arm.

- In all patients: Time from randomisation to the date of death or to the date of censoring for censored patients, regardless of treatment.

Subgroup analyses

Subgroup analyses will be conducted comparing OS between treatments in the following subgroups of the FAS:

- Age at randomisation (<65 vs ≥65 years of age)
- Sex (male vs female)
- Smoking status (smoker vs non-smoker).
- Stage of disease at study entry (Stage III A vs Stage III B).
- Histology (squamous vs all other).
- Best response to prior anticancer therapy (CR, PR, SD).
- Type of chemotherapy 1 (gemcitabine-based vs non- gemcitabine -based).
- Type of chemotherapy 2 (cisplatin-based vs carboplatin-based).
- Time from last dose of radiation to randomization (<14 days, ≥14 days).
- WHO performance status at baseline (normal activity [PSTAT=0] vs restricted activity [PSTAT=1]).
    - This will be determined from the response to "Performance status" (PSTAT module) on the eCRF at screening. Patients with a missing performance status will be included in the 'restricted activity' category.
- Region (Asia, Europe, South America vs North America).
    - This will be determined from the centre number (CENTRE). If there are less than 20 events in the "South America" category, these patients will be combined with those in North America.
- Race (White, Black/African-American, Asian, Other [Native Hawaiian/Pacific Islander or American Indian/Alaska Native or Others]).
    - This will be determined from the response to "Race" (DEM module) on the eCRF at screening.
- PD-L1 status (≥ 25% vs <25%).
    - Only samples with sample date on or before first dose should be used

  - for multiple evaluable samples with sample date on or before the first dose, the sample with the highest PD-L1 % tumour membrane staining result should be used

- EGFR (positive vs negative)

The subgroup analyses will be based on values recorded on the eCRF.

Other baseline variables may also be assessed if there is clinical justification or an imbalance is observed between the treatment arms. The purpose of the subgroup analyses is to assess the consistency of treatment effect across expected prognostic factors.

No adjustment to the significance level for testing will be made since all these subgroup analyses will be considered exploratory and may only be supportive of the primary analysis of OS.

For each subgroup, the HR and 95% CI will be calculated from an unstratified Cox proportional hazards model with treatment as only covariate. The Cox models will be fitted using SAS® PROC PHREG with the Efron method to control for ties, using the by statement to obtain HR and 95% CI for each subgroup level separately.

These hazard ratios and associated two-sided 95% CIs will be summarised and presented on a forest plot, along with the results of the overall primary analysis.

If there are too few events available for a meaningful analysis of a particular subgroup (it is not considered appropriate to present analyses where there are less than 20 events in a subgroup), the relationship between that subgroup and OS will not be formally analysed. In this case, only descriptive summaries will be provided.

Effect of covariates on the HR estimate

Cox proportional hazards modelling will be employed to assess the effect of covariates on the HR estimate for the primary treatment comparison. A model will be constructed, containing treatment and the stratification factors alone, to ensure any output from the Cox modelling is likely to be consistent with the results of the stratified log-rank test.

The result from the initial model and the model containing additional covariates will be presented.

Additional covariates for this model will be age at randomization, sex, smoking status, stage of disease at study entry, histology, best response to prior anticancer therapy, WHO performance status, region and race.

The model will include the effect regardless of whether the inclusion of effect significantly improves the fit of the model providing there is enough data to make them meaningful.

Consistency of treatment effect between subgroups

Interactions between treatment and stratification factors will also be tested to rule out any qualitative interaction using the approach of Gail and Simon 1985.

Impact of switching (crossover outside of this study) to immunotherapies (or other potentially active investigational agents) on OS analyses

Exploratory analyses of OS adjusting for the impact of subsequent immunotherapy or other treatment may be performed, if a sufficient proportion of patients switch.

Methods such as Inverse Probability of Censoring Weighting (Robins 1993), Rank Preserving Structural Failure Time (Robins and Tsiatis 1991), and other methods in development will be explored as appropriate. The decision to adjust and the final choice of methods will be based on a blinded review of the data and the plausibility of the underlying assumptions. Additional supportive analysis by censoring the treatments switchers at the time of initiation of subsequent therapy will also be conducted. These analyses are intended to further support the OS benefit and reimbursement appraisals.

Baseline and time-dependent characteristics will be explored, and summaries of baseline characteristics will be provided by treatment arm, splitting between those that have and haven't switched at the time of the analyses.

Subsequent therapies received after discontinuation of treatment will be summarised and listed by treatment group. Patients who subsequently received an immunotherapy agent or entered an immunotherapy trial will be summarised and listed by treatment arm according to line of subsequent therapy, i.e. immediately after immunotherapy or as a later line.

4.2.1.2 Progression free survival

PFS based upon the programmatically derived RECIST outcome using the BICR data (including all scans regardless of whether they were scheduled or not) will be analysed by stratified log-rank tests, the same methodology as described for the OS analyses.

The primary analysis of the co-primary endpoint, PFS, will occur when it is expected that 458 PFS events have occurred (65% maturity). PFS based upon the BICR data will be analyzed using a stratified log-rank test adjusting for the same factors as for OS. The effect of treatment will be estimated by the HR together with its corresponding (1-alpha adjusted)% CI and p-value for the full ITT population. Kaplan-Meier plots of PFS will be presented by treatment arm. Summaries of the number and percentage of patients experiencing a PFS event, and the type of event (RECIST 1.1 or death) will be provided along with median PFS for each treatment.

The assumption of proportionality will be assessed in the same way as for OS. The analysis will be based on the programmatically derived PFS using BICR data.

Supportive summaries/graphs

The treatment status at progression of patients at the time of analysis will be summarised. This will include the number (%) of patients who were on treatment at the time of progression,

the number (%) of patients who discontinued study treatment prior to progression, the number (%) of patients who have not progressed and were on treatment or discontinued treatment. This will also provide distribution of number of days prior to progression for the patients who have discontinued treatment.

The number of patients prematurely censored will be summarised by treatment arm. A patient would be defined as prematurely censored if they had not progressed (or died in the absence of progression) and the latest scan prior to DCO was more than one scheduled tumour assessment interval plus 2 weeks (10 weeks if time period between randomisation and DCO for that patient is 48 weeks or less; 14 weeks otherwise) prior to the DCO date.

Additionally, summary statistics will be given for the number of days from censoring to DCO for all censored patients.

A summary of the duration of follow-up will be summarised using median time from randomisation to date of censoring (date last known to be non-progressor) in censored (not progressed) patients only, presented by treatment group.

Additionally, summary statistics for the number of weeks between the time of progression and the last RECIST assessment prior to progression will be presented for each treatment group.

Summaries of the number and percentage of patients who miss two or more consecutive RECIST assessments will be presented for each treatment group.

All of the collected RECIST 1.1 data will be listed for all randomised patients. In addition, a summary of new lesions (i.e., sites of new lesions) will be produced.

Sensitivity Analyses

The following sensitivity analyses will be performed:

- Evaluation-Time bias

Sensitivity analyses will be performed to assess possible evaluation-time bias that may be introduced if scans are not performed at the protocol-scheduled timepoints. The midpoint between the time of progression and the previous RECIST assessment will be analysed using a stratified log-rank test, as described for the co-primary analysis of PFS. For patients whose death was treated as PFS event, the date of death will be used to derive the PFS time used in the analysis. This approach has been shown to be robust to even highly asymmetric assessment schedules (Sun and Chen 2010). To support this analysis, the mean of subject-level average inter-assessment times will be tabulated for each treatment. This approach will use the BICR data.

- Attrition bias

Attrition bias will be assessed by repeating the co-primary PFS analysis except that the actual PFS event times, rather than the censored times, of patients who progressed or died in the absence of progression immediately following two, or more, tumour assessments will be included. In addition, patients who take subsequent therapy prior to their last evaluable RECIST assessment or progression or death will be censored at their last assessment prior to taking the subsequent therapy. This analysis will be supported by a Kaplan-Meier plot of the time to censoring where the censoring indicator of the PFS analysis is reversed. This approach will use the BICR data.

- Ascertainment bias

Ascertainment bias will be assessed by analysing the site investigator data. The stratified log-rank test will be repeated on PFS using the site investigator data based upon RECIST. If there is an important discrepancy between the primary analysis using the BICR assessments and this sensitivity analysis using investigator assessments, then the proportion of patients with site but no central confirmation of progression will be summarised; such patients have the potential to introduce bias in the central review due to informative censoring. An approach that imputes an event at the next visit in the central review analysis may help inform the most likely HR value (Fleischer et al 2011), but only if an important discrepancy exists.

Disagreements between investigator and central reviews of RECIST progression will be presented for each treatment group. The summary will include the early discrepancy rate which is the frequency of central review declared progressions before the investigator review as a proportion of all central review progressions and the late discrepancy rate which is the frequency of central review declared progressions after the investigator review as a proportion of all discrepancies.

- Confirmation bias

An additional sensitivity analysis using the BICR data will be performed to determine the effect of confirmation of progression. The stratified log-rank test will be repeated on PFS using BICR data based upon RECIST modified for confirmation of progression. The HR and CI will be presented.

Subgroup/Additional Analysis

An exploratory analysis of PFS using the irRECIST 1.1 data provided by the BICR will be performed on the BICR analysis set. The stratified log-rank test will be repeated on PFS using the BICR based upon the irRECIST 1.1 data. The HR and CI will be presented.

Subgroup analyses and a forest plot will be generated comparing PFS between treatments in the same way as previously specified for OS. Unless there is a marked difference between the results of the statistical analyses of the PFS from the BICR data and that of the site investigator, this will only be performed upon the PFS endpoint using the BICR data based upon RECIST. No adjustment to the significance level for testing will be made since all these

subgroup and sensitivity analyses will be considered supportive of the primary analysis of PFS.

A forest plot illustrating the hazard ratio and the corresponding 95% confidence interval will be provided to compare the primary and sensitivity analyses of progression free survival.

The effect of covariates upon the HR estimate and the consistency of treatment effect between subgroups will be analysed for PFS using the same methods as those described for OS.

A further analysis of PFS (using investigator assessed RECIST) may be performed at the time of the OS analyses, if requested by health authorities.

Time to first subsequent therapy or death (TFST)

For supportive purposes, the time to the start of subsequent therapy will be analysed using the same methodology and model as that used for the co-primary analysis of PFS. The HR for the treatment effect together with its 95% CI will be presented. In addition, a Kaplan-Meier plot of the time to the start of subsequent therapy will be presented by treatment arm and the time between progression and starting subsequent therapy will be assessed based upon the investigator data defined date of progression. This will be summarised per treatment arm but no formal comparisons will be made. No multiplicity adjustment will be applied as this is viewed as a supportive endpoint.

In patients who received a subsequent cancer therapy, a summary table of first subsequent cancer therapies by treatment arm will be provided.

4.2.2 Secondary efficacy endpoints

4.2.2.1 Proportion of patients alive at 24 months (OS24)

OS24 will be analysed (using the Kaplan-Meier curve) and presented by treatment arm. For each treatment arm, the survival rate at 24 months based on Kaplan-Meier method will be presented, along with its 95% confidence interval. The computation of the confidence interval will be based on a log(-log(.)) transformation.

For the comparison between treatments (sub-study B only), the test will be based on the method described in Klein 2007 (Klein et al 2007). The test statistic and its variance estimate are as follows:

- test statistic = $\ln \frac{\ln \hat{S}_1(t)}{\ln \hat{S}_2(t)}$
- Variance estimate = $\frac{\hat{\sigma}_1(t)^2}{\ln^2 S_1(t)} + \frac{\hat{\sigma}_2(t)^2}{\ln^2 S_2(t)}$

where $\hat{\sigma}_i(t)^2 = \sum_{t_i \le t} \frac{d_i}{n_i(n_i - d_i)}$ is the variance derived from Greenwood's formula *S(t)*and can be estimated from standard software packages, where d_i and n_i refer to the number of deaths and patients at risk for each risk set.

The z-statistic is then calculated as: $\frac{\text{test statistic}}{\sqrt{\text{variance estimate}}}$

For stratified analysis, the test statistic and its variance estimate in each strata will be combined by weighting inversely proportionately according to each within stratum variance (Whitehead and Whitehead 1991). A Z-test will be performed and the p-value from the test will be presented.

4.2.2.2 Objective response rate

The ORR will be based on the programmatically derived RECIST outcome using the BICR data. The treatment comparison of ORR will be based on a Fisher's exact test using mid p-values. A binary response variable for ORR will be used for the analysis with the categories of CR and PR vs SD, PD and NE. The results of the analysis will be presented in terms of the proportion of responders and a p-value.

The mid-p-value modification of the Fisher exact test amounts to subtracting half of the probability of the observed table from Fisher's p-value.

$$\text{Fisher's exact test mid p-value} = \text{Two sided p-value} - \frac{\text{Table probability}}{2}$$

This analysis of ORR will be repeated using the results of the programmatically derived outcome using the site investigator data based upon RECIST as a sensitivity analysis to confirm the results of the primary analysis based on the site investigator tumour data. An additional sensitivity analysis will be performed on programmatically derived outcomes using BICR tumour data based upon RECIST modified for confirmation of progression to determine if there is any difference when using progression confirmation rules. Finally, an exploratory analysis of ORR using the irRECIST 1.1 data obtained from the BICR will be performed where the above analysis will be repeated.

Summaries will be produced that present the number and percentage of patients with a tumour response (CR/PR) based upon the number of patients with measurable disease at baseline per BICR or investigator as appropriate (see Section 3.2.2.2). For each treatment arm, best objective response (BoR) will be summarised by n (%) for each category (CR, PR, ND, SD, PD, NE). This will be produced for the BICR of RECIST only. No formal statistical analyses are planned for BoR.

4.2.2.3 Duration of response

Descriptive data will be provided for the DoR in responding patients (ie median duration of response and 95% CIs) by treatment arm, including the associated Kaplan-Meier curves (without any formal comparison of treatment arms or p-value attached).

4.2.2.4 Proportion of patients alive and progression free at 12 months (APF12)

The proportion of patients alive and progression free at 12 months (i.e. at study day 366) will be summarised (using the Kaplan-Meier curve) and presented by treatment arm. For each treatment arm, the APF12 based on Kaplan-Meier method will be presented, along with its 95% confidence interval. The computation of the confidence interval will be based on a log(-log(.)) transformation.

4.2.2.5 Proportion of patients alive and progression free at 18 months (APF18)

The proportion of patients alive and progression free at 18 months (i.e. at study day 549) will be summarised (using the Kaplan-Meier curve) and presented by treatment arm. For each treatment arm, the APF18 based on Kaplan-Meier method will be presented, along with its 95% confidence interval. The computation of the confidence interval will be based on a log(-log(.)) transformation.

4.2.2.6 Time from randomisation to second progression (PFS2)

Time from randomisation to second progression (PFS2) will be analysed using the same methods as outlined for the analysis of PFS and adjusting for the same set of covariates, but no subgroup analysis will be performed. Medians and Kaplan–Meier plots will be presented to support the analysis. The sensitivity analysis outlined in Section 4.2.1.2 will not be repeated for PFS2 with the exception of a Kaplan-Meier plot of the time to censoring where the censoring indicator of PFS2 is reversed.

The number and percentage of subjects experiencing a PFS2 event and the type of progression (objective progression by RECIST, symptomatic progression, new or worsening of soft tissue/visceral or bone metastases or other) will also be summarised by treatment arm.

Time to second subsequent therapy or death (TSST)

For supportive purposes, the time to the start of second subsequent therapy will be analysed using the same methodology and model as that used for the analysis of PFS2. The HR for the treatment effect together with its 95% CI will be presented. In addition, a Kaplan-Meier plot of the time to the start of second subsequent therapy will be presented by treatment arm and the time between progression and starting second subsequent therapy will be assessed based upon the site investigator data defined date of progression. This will be summarised per treatment arm but no formal comparisons will be made. No multiplicity adjustment will be applied as this is viewed as a supportive endpoint.

4.2.2.7 Time to death or distant metastasis

TTDM will be analysed using identical methods as outlined for the analysis of PFS and adjusting for the same set of covariates, but no subgroup analysis will be performed. Medians and Kaplan-Meier plots will be presented to support the analysis. The sensitivity analysis outlined in Section 4.2.1.2 will not be repeated for TTDM with the exception of a Kaplan Meier plot of the time to censoring where the censoring indicator of TTDM is reversed.

4.2.2.8 Change in tumour size

The absolute values and percentage change in TL tumour size from baseline will be summarized using descriptive statistics and presented at each timepoint by treatment arm. The best change in TL tumour size from baseline, (where best change in TL size is the maximum reduction from baseline or the minimum increase from baseline in the absence of a reduction) will also be summarised and presented for each treatment arm.

Tumour size will also be presented graphically using waterfall plots for each treatment arm, to present each subject's best percentage change in tumour size as a separate bar, with the bars ordered from the largest increase to the largest decrease. Reference lines at the +20% and –30% change in tumour size levels will be added to the plots, which correspond with the definitions of progression and 'partial' response respectively. On each of the waterfall plots the PD-L1 status (positive vs negative), the histology categorization (squamous vs other) and the disease status at study entry (IIIA vs IIIB) of each patient will be indicated as appropriate. Additional waterfall plots showing percentage change in tumour size at specific timepoints may be produced if it is felt that these are warranted to provide greater clarity.

The above outputs will be programmed on data based upon BICR RECIST assessments. They will be repeated for the irRC (irRECIST 1.1) data obtained from BICR where tumour size is defined as the sum of the diameters of TLs and the new measured lesions.

4.2.3 Secondary patient reported outcome endpoints

The PRO endpoints that have been identified as primary are EORTC QLQ-C30 time to HRQoL deterioration for global health status and LC13 time to symptom deterioration for each of dyspnoea, cough, haemoptysis, and chest pain. These are not part of the main multiple testing procedure and as supportive endpoints will need a Bonferroni adjustment to the significance level to aid interpretation. Therefore, these 5 endpoints will be tested at a 1% significance level and 99% CIs will be produced.

The other time to symptom deterioration endpoints will be tested at a 5% significance level and 95% CIs will be produced.

4.2.3.1 EORTC QLQ-C30

Time to symptom deterioration will be analysed for each of the 3 symptom scales (fatigue, pain, nausea/vomiting) and the 5 individual symptom items (dyspnoea, insomnia, appetite loss, constipation, diarrhoea).

Time to HRQoL/function deterioration will be analysed for the 5 function scales (physical, role, emotional, cognitive, and social) and global health status/HRQoL. For treatment comparison, a stratified log-rank test will be used, as described for the primary analysis of OS.

The HR and 95% CI for each scale/item will be presented graphically on a forest plot.

A summary of the symptom improvement rate for each of the 3 symptom scales and the 5 individual symptom items will be produced. Similarly, a summary of HR QoL/function improvement rate for each of the 5 function scales (physical, role, emotional, cognitive, and social) and global health status/HR QoL will be produced. Symptom improvement rate and HR QoL/function improvement rate will be analysed by comparing between treatment arms using a logistic regression model adjusting for the same factors as the co-primary endpoints. The results of the analysis will be presented in terms of an odds ratio (an odds ratio greater than 1 will favour treatment A) together with its associated profile likelihood 95% CI (e.g. using the option 'LRCI' in SAS procedure GENMOD) and p-value (based on twice the change in log-likelihood resulting from the addition of a treatment factor to the model). The odds ratio and 95% CI for each scale/item will be presented graphically on a forest plot. If there are very few responses in one treatment arm, a Fisher's exact test will be considered.

For each of the 3 symptom scales (fatigue, pain, nausea/vomiting), 5 individual symptom items (dyspnoea, insomnia, appetite loss, constipation, diarrhoea), 5 functional scales (physical, role, emotional, cognitive, and social), and global health status/HR QoL, time to deterioration will be presented using a Kaplan-Meier plot. Summaries of the number and percentage of patients experiencing a clinically meaningful deterioration or death, and the median time to deterioration will also be provided for each treatment arm.

Summaries of original and change from baseline values of each symptom scale/item, the global HRQoL score and each functional domain will be reported by visit for each treatment arm. Graphical presentations may also be produced as appropriate. Summaries of the number and percentage of patients in each response category at each visit for each ordinal item (in terms of the proportion of patients in the categories of improvement, no change, and deterioration as defined in Section 3.3.1) will also be produced for each treatment arm.

A summary of compliance rate and evaluability rate will be provided for each treatment arm, by assessment time point and also for overall.

4.2.3.2 EORTC-QLQ-LC13

Time to symptom deterioration for each of the 6 individual symptoms (dyspnoea, cough, haemoptysis, chest pain, arm/shoulder pain, other pain) will be compared between treatment arms using a stratified log-rank test as described for the primary analysis of OS.

The HR and 95% CI for each scale/item will be presented graphically on a forest plot.

For each of the 6 symptoms items in LC13, time to deterioration in symptoms will be presented using a Kaplan-Meier plot. Summaries of the number and percentage of patients experiencing a clinically meaningful deterioration or death, and the median time to deterioration will also be provided for each treatment arm.

A summary of the symptom improvement rate for each of the 6 individual symptom items will be produced. The symptom improvement rate will be compared between treatment arms using a logistic regression model as described for EORTC-QLQ-C30. The odds ratio and 95% CI for each symptom will be presented graphically on a forest plot. If there are very few responses in one treatment arm, a Fisher's exact test will be considered.

Summaries of original and change from baseline values of each symptom (dyspnoea, cough, haemoptysis, chest pain, arm/shoulder pain, other pain) and each treatment-related side effect (sore mouth, dysphagia, peripheral neuropathy and alopecia) will be reported by visit for each treatment arm. Graphical presentations may also be produced as appropriate. Summaries of the number and percentage of patients in each response category at each visit for each ordinal symptom item (in terms of the proportion of patients in the categories of improvement, no change, and deterioration as defined in Section 3.3.2 will also be produced for each treatment arm.

A summary of compliance rate and evaluability rate will be provided for each treatment arm, by assessment time point and also for overall.

4.2.3.3 Mixed models repeated measures of change from baseline in PRO symptoms

In addition to the time to deterioration endpoints listed above the following longitudinal endpoints are of interest: EORTC QLQ-C30 global health status, physical functioning, fatigue and appetite loss; LC13 dyspnoea, cough and chest pain. These are not part of the main multiple testing procedure and are considered a separate set of PRO endpoints from the time to deterioration endpoints listed above. A Bonferroni adjustment to the significance level will be applied to the tests described below to control the overall Type I error at the 5% level.

Change from baseline in these pre-specified the PRO symptom scores of dyspnea, cough, chest pain, fatigue and appetite loss, global health status and physical functioning will be analysed using a mixed model for repeated measures (MMRM) analysis making use of all data from baseline up to 12 months. The analysis will be to compare the average treatment effect from the point of randomisation until PD or 12 months (whichever is earlier) unless there is excessive missing data (defined as >75% missing data). It is acknowledged that patients will discontinue treatment at different timepoints during the study and that this is an important time with regards to symptoms and HRQoL data collection. To account for this, and in order to include the discontinuation and follow up visits, a generic visit variable will be derived for each subject in order that the average treatment effect can be analyzed using the above method. Each visit will be assigned a sequential number. The time from randomization

to each of these will be derived in order to select only those visits occurring within the first 12 months of randomization or until PD.

As an example, say a patient X attends the first 4 scheduled visits of a 4-weekly schedule and then discontinues treatment, whilst patient Y discontinues treatment after the first scheduled visit, the first 6 generic visits would be as follows:

Generic visit	Study Day	
	Patient X	Patient Y
Baseline	Baseline	Baseline
1	29	28
2	57	50 (discontinuation)
3	85	85
4	113	113
5	130 (discontinuation)	141
6	169	169

The MMRM model will include the fixed, categorical effects of treatment, visit, and treatment-by-visit interaction, age at randomisation (<65 vs ≥65 years of age), sex (male vs female), smoking history (smoker vs non-smoker as well as the continuous fixed covariate of baseline score and the baseline score-by-visit interaction. Restricted maximum likelihood (REML) estimation will be used. An overall adjusted mean estimate will be derived that will estimate the average treatment effect over visits giving each visit equal weight. For this overall treatment comparison, adjusted mean estimates per treatment group and corresponding 95% CIs will be presented along with an estimate of the treatment difference, 95% CI and p-value

An unstructured covariance matrix will be used to model the within-subject error and the Kenward-Roger approximation will be used to estimate the degrees of freedom. If the fit of the unstructured covariance structure fails to converge, the following covariance structures will be tried in order until convergence is reached: toeplitz with heterogeneity, autoregressive with heterogeneity, toeplitz, autoregressive and compound symmetry.

Multiple imputation techniques for missing values may be considered to explore the robustness of any treatment effect.

An effect size estimate to interpret the magnitude of the effect and potential therapeutic benefit will be further specified in the PAP.

4.2.4 Safety data

Safety and tolerability data will be presented by treatment arm using the safety population. Safety summaries will be descriptive only. No formal statistical analyses will be performed on the safety variables.

Any safety summaries examining re-treatment of study medication will be produced separately as needed.

The following sections describe the planned safety summaries for AEs, vital signs, laboratory parameters, ECG and WHO performance status. However, additional safety summaries (not specified in this SAP) may need to be produced to aid interpretation of the safety data.

4.2.4.1 Adverse events

All AEs, both in terms of current MedDRA preferred term and CTCAE grade, will be summarised descriptively by count (n) and percentage (%) for each treatment group. Any AE occurring before randomized treatment (i.e. before the administration of the first infusion on Study Day 1) will be included in the AE listings, but will not be included in the summary tables (unless otherwise stated). These will be referred to as 'pre-treatment'.

AEs observed up until 90 days following discontinuation of the study treatment (i.e., the last dose of randomised treatment) or until the initiation of the first subsequent anti-cancer therapy (including radiotherapy, with the exception of palliative radiotherapy) following discontinuation of study treatment (whichever occurs first) will be used for reporting of all of the AE summary tables. This will more accurately depict AEs attributable to study treatment only as a number of AEs up to 90 days following discontinuation of the study treatment are likely to be attributable to subsequent therapy.

However, to assess the longer term toxicity profile, all of the AE summaries will also be produced containing AEs observed up until 90 days following discontinuation of the study treatment (ie without taking subsequent therapy into account).

All reported AEs will be listed along with the date of onset, date of resolution (if AE is resolved) and investigator's assessment of severity and relationship to study drug. Frequencies and percentages of patients reporting each preferred term will be presented (i.e. multiple events per patient will not be accounted for apart from on the episode level summaries).

Summary information (the number and percent of patients by system organ class (SOC) and PT separated by treatment group) will be tabulated for:

- All AEs
- All AEs causally related to study medication (as determined by the reporting investigator)

- AEs with CTCAE grade 3 or higher
- AEs with CTCAE grade 3 or higher, causally related to study medication (as determined by the reporting investigator)
- Most common AEs
- Most common AEs with CTCAE grade 3 or higher
- AEs with outcome of death
- AEs with outcome of death causally related to study medication (as determined by the reporting investigator)
- All SAEs
- All SAEs causally related to study medication (as determined by the reporting investigator)
- AEs leading to discontinuation of study medication
- AEs leading to discontinuation of study medication, causally related to study medication (as determined by the reporting investigator)
- AEs leading to dose interruption of study medication
- Other significant AEs
- Immune mediated AEs (as determined by the reporting investigator)
- Infusion reaction AEs (as determined by the reporting investigator)

An overall summary of the number and percentage of patients in each category will be presented, as will an overall summary of the number of episodes in each category. In addition, a truncated AE table of most common AEs and another table showing most common AEs with CTCAE grade 3 or higher, showing all events that occur in at least 5% of patients overall will be summarised by preferred term, by decreasing frequency in the total column (although the total column may not be displayed in the AE tables). This cut-off may be modified after review of the data. When applying a cut-off (i.e., x %), the raw percentage should be compared to the cut-off, no rounding should be applied first (i.e., an AE with frequency of 4.9% will not appear if a cut-off is 5%). Summary statistics showing the time to onset and the duration of the first AE may also be presented as appropriate.

Each AE event rate (per 100 patient years) will also be summarised by preferred term within each SOC for the output summarising all AEs. For each preferred term, the event rate is defined as the number of patients with that AE divided by the total drug exposure of patients

at risk of AE divided by the total drug exposure of patients and then multiplied by 365.25 x 100 to present in terms of per 100 patient years.

Summaries of the number and percentage of patients will be provided by maximum reported CTCAE grade, SOC, PT and treatment group.

Fluctuations observed in CTCAE grades during study will be listed for those AEs which are CTCAE ≥ 3.

In addition, all AEs will be listed.

Deaths

Two summaries of all deaths will be provided with number and percentage of patients by treatment group, categorised as:

- Total number of deaths (regardless of date of death)
- Related to disease under investigation only
- AE outcome of death only and onset date prior to initiation of subsequent anti-cancer therapy
- AE outcome of death only and onset date falling after 90 days following last dose of study medication or initiation of subsequent anti-cancer therapy (whichever is earlier)
- Both related to disease under investigation and with AE outcome=death and onset date prior to initiation of subsequent anti-cancer therapy
- Death related to disease under investigation and AE with outcome=death > 90 days after last dose of study medication or ≥ date of subsequent therapy, whichever occurs first
- Deaths > 90 days after last dose of study medication or ≥ date of subsequent therapy (whichever occurs first), unrelated to AE or disease under investigation
- Patients with unknown reason for death.
- Other deaths

This summary will be repeated for all deaths within 90 days of last dose of study medication.

Adverse events of special interest

PTs used to identify adverse events of special interest (AESI), as defined in Section 3.4.1, will be listed before DBL and documented in the Study Master File.

Grouped summary tables of certain MedDRA PTs will be produced. For each 'grouped' term, the number (%) of patients experiencing any of the specified terms will be presented by maximum CTCAE grade. Additional summaries will include Time to Onset of first CTCAE grade 3 or higher. Time to onset of first AE for each grouped term and preferred term within it will also be produced. Groupings will be based on preferred terms provided by the medical team prior to DBL, and a listing of the preferred terms in each grouping will be provided.

Additional summaries of the above-mentioned grouped AE categories will include number (%) of patients who have:

- At least one AESI presented by outcome
- At least one AESI causally related to study medication
- At least one AESI leading to discontinuation of study medication

A summary of total duration (days) of selected AESI will be provided. Additional information to be summarised will include number of patients with an event (n %), number of episodes of event (n), median total duration of event (days, range) and median duration of CTCAE grades 1, 2, 3, 4 (days, range).

Additionally, there will be several summaries of AESIs requiring concomitant treatment, and particularly the relationship of AESIs to the use of immunosuppressive agents (ie, depicting which AESI triggered immunosuppressive use) and, separately, to the use of immunosuppressive agents at high doses."

4.2.4.2 Laboratory assessments

Data obtained up until the 90 days following discontinuation of study treatment or until the initiation of the first subsequent anti-cancer therapy (including radiotherapy, with the exception of palliative radiotherapy) following discontinuation of study treatment (whichever occurs first) will be used for reporting. This will more accurately depict laboratory toxicities attributable to study treatment only as a number of toxicities up to 90 days following discontinuation of the study treatment are likely to be attributable to subsequent therapy.

However, to assess the longer term toxicity profile, some summaries of laboratory data may be produced containing data collected up until 90 days following discontinuation of the study treatment (i.e., without taking subsequent therapy into account).

Any data post 90 days after the last dose of the study treatment will not be summarised.

Data summaries will be provided in preferred units.

Scatter plots (shift plots) of baseline to maximum value/minimum value (as appropriate) on treatment (i.e. on-treatment is defined as data collected between the start of treatment and the relevant follow-up period following the last dose of study treatment) may be produced for certain parameters if warranted after data review.

Box-plots of absolute values by week, and box-plots of change from baseline by week, may be presented for certain parameters if warranted after data review. For continuous laboratory assessments absolute value and change from baseline will be summarised using descriptive statistics at each scheduled assessment time by actual treatment group.

Shift tables for laboratory values by worst CTCAE grade will be produced, and for specific parameters separate shift tables indicating hyper- and hypo- directionality of change will be produced. The laboratory parameters for which CTCAE grade shift outputs will be produced are:

- Haematology: Haemoglobin; Leukocytes; Lymphocytes (count, absolute); Neutrophils (count, absolute);Platelets

- Clinical chemistry: ALT, AST, ALP, Total Bilirubin, Albumin, Magnesium – hypo and – hyper, Sodium – hypo and – hyper, Potassium – hypo and – hyper, Corrected Calcium – hypo and – hyper, Glucose – hypo and – hyper, Creatinine.

Additional summaries will include a shift table for urinalysis (Bilirubin, Blood, Glucose, Ketones, Protein) comparing baseline value to maximum on-treatment value.

Liver Enzyme Elevations and Hy's law

The following summaries will include the number (%) of patients who have:

- Elevated ALT, AST, and Total bilirubin during the study

  – ALT ≥ 3x –≤ 5x, > 5x – ≤8x, > 8x - ≤ 10x, >10x - ≤ 20x, and >20x Upper Limit of Normal (ULN) during the study

  – AST ≥ 3x–≤ 5x, > 5x – ≤8x, > 8x - ≤ 10x, >10x - ≤ 20x, and >20x ULN during the study

  – Total bilirubin ≥2x-≤3x, >3x-≤5x, >5x ULN during the study

  – ALT or AST ≥3x-≤5x, >5x - ≤8x, >8x - ≤ 10x, >10x - ≤ 20x, >20x ULN during the study

  – ALT or AST ≥3x ULN and Total bilirubin ≥2x ULN during the study (Potential Hy's law): The onset date of ALT or AST elevation should be prior to or on the date of Total Bilirubin elevation

Liver biochemistry test results over time for patients with elevated ALT or AST (ie ≥ 3x ULN), and elevated Total Bilirubin (ie ≥ 2x ULN) (at any time) will be plotted. Individual patient data where ALT or AST (ie ≥ 3x ULN) plus Total Bilirubin (ie ≥ 2x ULN) are elevated at any time will be listed also.

Plots of ALT and AST vs. Total Bilirubin by treatment group will also be produced with reference lines at 3×ULN for ALT, AST, and 2×ULN for Total Bilirubin. In each plot, Total Bilirubin will be in the vertical axis.

4.2.4.3 ECGs

ECG data obtained up until the safety follow-up will be included in the summary tables.

Overall evaluation of ECG is collected at each visit in terms of normal or abnormal, and the relevance of the abnormality is termed as “clinically significant” or “not clinically significant”. A shift table of baseline evaluation to worst evaluation on-treatment will be produced.

4.2.4.4 Vital signs

Vital signs data obtained up until the 30 day safety follow-up visit will be included in the summary tables.

Box plots for absolute values and change from baseline by week may be presented for certain vital signs parameters if warranted after data review.

Vital signs (systolic blood pressure (SBP), diastolic blood pressure (DBP), pulse rate, temperature, respiratory rate and weight) will be summarised over time in terms of absolute values and change from baseline at each scheduled measurement by actual treatment group

4.2.4.5 Physical examination

Individual physical examination data will not be summarised.

4.2.4.6 Other safety data

Data from positive pregnancy tests will not be summarised.

4.2.5 WHO performance status

All WHO performance status data will be summarised over time for the ITT population.

4.2.6 PK analysis

MEDI4736 concentration data will be listed, and a summary will be provided for all evaluable patients. These outputs will be produced by AstraZeneca/MedImmune Clinical Pharmacology group or designee.

4.2.7 Immunogenicity analysis

Immunogenicity results will be listed and a summary will be provided of the number and percentage of patients who develop detectable anti-MEDI4736 antibodies based on the safety population. The immunogenicity titre and neutralising ADA data will be listed for samples confirmed positive for the presence of anti-MEDI4736 antibodies. The effect of immunogenicity on PK, PDx, efficacy and safety will be evaluated if data allow. These outputs will be produced by AstraZeneca/MedImmune Clinical Pharmacology group or designee.

4.2.8 PK/PDx relationships

If the data are suitable, the relationship between MEDI4736 PK exposure and efficacy/safety parameters may be investigated graphically or using appropriate data modelling approach. These outputs will be produced by AstraZeneca/MedImmune Clinical Pharmacology group or designee.

4.2.9 Exploratory endpoints

4.2.9.1 EQ-5D-5L

The change from baseline in health state utility values and the visual analogue scale will be compared between treatment arms at each visit using a mixed model repeated measures analysis, which adjusts for the same factors as the primary analysis and also the baseline health state utility value/visual analogue scale as appropriate. Adjusted mean differences between treatments and 95% CIs from these analyses will be presented, but, as this analysis is exploratory in nature, p-values will not be calculated. Further detail of this analysis will be provided in the payer analysis plan.

Descriptive statistics will be reported for health state domain (e.g., proportion in each domain) and the visual analogue scale by visit, as well as the change in the visual analogue scale value and the derived utility index value from baseline.

To support future economic evaluations, additional appropriate analyses may be undertaken, for example, mean health state utility pre- and post-treatment, and pre- and post-progression. These exploratory analyses may be carried out to support health authority appraisals and will consequently not be reported in the CSR. Further detail will be provided in the payer analysis plan.

A summary of compliance rate and evaluability rate will be provided for each treatment arm, by assessment time point and also for overall as appropriate.

4.2.9.2 Health resource use

An exploratory health economic analysis of hospital episodes including the number of hospitalisations, type of contact (hospitalisation, outpatient, day case), reason, total length of hospital stay and total length of ICU stay will be undertaken to examine the impact of disease and treatment on resource use to primarily support the economic evaluation of MEDI4736. This will include providing descriptive statistics as appropriate, including means, median,

ranges or frequencies and percentages and will be summarised by randomised treatment arm. These simple summaries will be included in the CSR. However, the exploratory health economic analysis to be performed will be described in the payer analysis plan.

4.2.10 Demographic and baseline characteristics data

The following will be summarised for all patients in the FAS (unless otherwise specified) by treatment group:

- Patient disposition (including screening failures and reason for screening failure)
- Important protocol deviations
- Inclusion in analysis populations
- Demographics (age, age group[<50, >=50-< 65, ≥ 65 years and additionally ≥ 75 years], sex, race and ethnicity)
- Patient characteristics at baseline (height, weight, weight group, body mass index (BMI) and BMI group)
- Patient recruitment by country and centre
- Previous disease-related treatment modalities
- Time from initial diagnosis to randomization
- Disease characteristics at baseline (WHO performance status, primary tumour location, histology type, tumour grade and overall disease classification, best response to previous therapy)
- Extent of disease at baseline
- TNM classification at baseline
- Medical history (past and current)
- Relevant surgical history (as appropriate)
- Physical examination at baseline
- Disallowed concomitant medications
- Allowed concomitant medications
- Post-discontinuation cancer therapy

- Nicotine use, categorised (never, current, former)
- Stratification factors as per IVRS and eCRF data

The AZ drug dictionary (AZDD) will be used for concomitant medication coding.

Patient disposition data will also be summarised at the time of OS analysis.

4.2.11 Treatment exposure and intensity

The following summaries related to study treatment will be produced for the safety analysis set by treatment group:

- Total exposure
- Actual exposure
- Reasons for dose delays/interruptions. Dose interruptions will be based on investigator initiated dosing decisions
- Number of infusions received
- RDI

For patients on study treatment at the time of the PFS and OS analysis, the DCO date will be used to calculate exposure.

5. INTERIM ANALYSES

5.1 Analysis Methods

Three interim analyses will be performed, one for PFS and two for OS.

The DCO for the interim analysis of PFS (the first interim analysis) will occur when it is expected that 367 PFS events have occurred (52% maturity, approximately 30 months after the first patient is randomised).

The DCO for the first interim analysis of OS (the second interim analysis) will occur when it is expected that 458 PFS events have occurred (65% maturity, approximately 36 months after the first patient is randomised). This is also the planned time for PFS final analysis. It is expected that approximately 285 death events (41% maturity) will be available for the interim OS analysis (assuming OS HR=0.73). The second interim analysis of OS (the third interim analysis) will be conducted when it is expected that 393 OS events have occurred (56% maturity, approximately 47 months after the first patient is randomised). The interim analyses will be assessed by an IDMC (further details are given in the IDMC charter). It is expected that recruitment will have completed prior to the results of these interim analyses being available

5.1.1 PFS interim analysis

Approximately 367 PFS events (52% maturity) will be available for the interim PFS analysis.

The Lan DeMets spending function that approximates an O'Brien Fleming approach will be used to account for multiplicity introduced by including an interim analysis for superiority (Lan and DeMets 1983).

The criterion for superiority is a statistically significant improvement in PFS at the interim analysis. If 80% of the PFS events required at the time of the primary PFS analysis is available at the time of the interim (i.e., 367/458 PFS events have occurred), the 2-sided significance level to be applied for the PFS interim analysis would be 1.04% and the 2-sided significance level to be applied for the primary PFS analysis would be 2.19%.

PFS will be analysed using a stratified log-rank test (see Section 4.2.1.2 for details). However subgroup, sensitivity, secondary analysis and the global interaction test will not be performed at the interim. The hazard ratio will be estimated with corresponding CI and p-value. The size of the CI will be determined based on the actual number of events included in the interim analysis.

The progression status of patients at the time of the interim analysis will be summarised. A Kaplan-Meier plot of PFS will be presented by randomised treatment group, along with median PFS. The number of progression events will be presented by randomised treatment group.

If the PFS results indicate superiority, then analyses of all other endpoints may also be performed. Patients would continue to be followed for PFS and survival until approximately 458 PFS events, when the primary PFS analysis and the first interim OS analysis would be performed.

If the PFS interim analysis result does not meet the criterion of stopping for superiority, then all patients will remain blinded and continue to be followed for PFS and OS.

The recommendations from the IDMC will not reveal the results of the analysis but will take the form of "Continue/Modify/Stop".

5.1.2 OS interim analyses

Approximately 285 and 393 death events (41% and 56% maturity) will be available for the first and second interim OS analysis, respectively (assuming OS HR=0.73).

The Lan DeMets spending function that approximates an O'Brien Fleming approach will be used to account for multiplicity introduced by including an interim analysis for superiority (Lan and DeMets 1983).

The criterion for superiority is a statistically significant improvement in OS at the interim analyses. If 58% or 80% of OS events required at the time of the primary OS analysis are available at the time of the interim (i.e., 285/491 or 393/491 events have occurred), the 2-sided significance level to be applied for the OS interim analysis would be 0.21% and 0.98%, respectively, and the 2-sided significance level to be applied for the primary OS analysis would be 2.17%.

OS will be analysed using a stratified log-rank test (see Section 4.2.1.2 for details). However subgroup, sensitivity, secondary analysis and the global interaction test will not be performed at the interims. The hazard ratio will be estimated with corresponding CI and p-value. The size of the CI will be determined based on the actual number of events included in the analysis.

The survival status and progression status of patients at the time of the interim OS analyses and primary PFS analysis will be summarised. Kaplan-Meier plots of OS and PFS will be presented by randomised treatment group, along with median OS and PFS. The number of death events and progression events will be presented by randomised treatment group.

If the PFS and/or OS results indicate superiority, then analyses of all other endpoints would be performed and the results of these analyses will form the basis for submissions for regulatory approval. Patients would continue to be followed for survival until approximately 491 patients have died, when an updated analysis would be performed.

If the PFS result is not statistically significant and/or the OS interim analyses results do not meet the criterion of stopping for superiority, then all patients will remain blinded and continue to be followed for survival.

The recommendations from the IDMC will not reveal the results of the analysis but will take the form of “Continue/Modify/Recommend Early Submission/Stop”.

5.2 Independent Data Monitoring Committee

This study will use an external IDMC to assess ongoing safety analyses as well as interim efficacy analyses for superiority based on PFS and OS and the primary analysis of PFS:

- The IDMC will review the safety data from approximately the first 75 patients, or approximately 3 months after randomisation of the first patient and then again 3 months later.
- The IDMC will then meet at least every 6 months thereafter to review safety data.
- Additional reviews of the safety data may be requested by the IDMC at additional points during the study.

This committee will be composed of therapeutic area experts and biostatisticians, who are not employed by AstraZeneca/MedImmune and do not have any major conflict of interest.

Following the reviews, the IDMC will recommend whether the study should continue unchanged, be stopped, or be modified in any way. Once the IDMC has reached a recommendation, a report will be provided to AstraZeneca/MedImmune. The report will include the recommendation and any potential protocol amendments, and will not contain any unblinding information. The final decision to modify or stop the study will sit with the sponsor.

In addition:

- The IDMC will review the efficacy data when 367 progression events (per BICR) have occurred, at approximately 30 months post-randomisation, at the time of the interim analysis of PFS.

- The IDMC will review the efficacy data when 458 progression events (per BICR) have occurred, at approximately 36 months post-randomisation at the time of the primary analysis of PFS and the first interim analysis of OS.

- The IDMC will review the efficacy data when 393 OS events have occurred, at approximately 47 months post-randomisation at the time of the second interim analysis of OS.

A separate IDMC charter will be developed, which will contain details of the IDMC members and clearly define the responsibilities of the IDMC.

The safety of all AstraZeneca/MedImmune clinical studies is closely monitored on an ongoing basis by AstraZeneca/MedImmune representatives in consultation with the Patient Safety Department. Issues identified will be addressed; this could involve, for instance, amendments to the clinical study protocol and letters to investigators.

6. CHANGES OF ANALYSIS FROM PROTOCOL

Table 9 Changes from the planned analysis

Section of SAP Affected (If applicable)	Change	Rationale
4.2.3 Patient Reported Outcomes	Add sensitivity analysis for primary PRO endpoints where RECIST progression counts as an event	To explore the effect of RECIST progression on the PRO endpoint analyses
2.1 Definition of Analysis Sets	Remove criteria that patients need to have post-dose data to be included in the safety analysis set	More widely used definition and to be consistent with other studies
Section 3.2.2.7 Time to death of Distant Metastasis	Clarified definition of Time to death or distant metastasis	Added further detail based on data will be collected
Section 4.2.1.1 Overall Survival	Removed global interaction test for subgroup analyses, subgroups will be tested using model only including treatment	More widely used definition and to be consistent with other studies
Section 4.2.2.1	Remove HR and added p-value to the presentation of OS12	p-value required for multiple testing procedure
Section 4.2.2.4/5	Change presentation of proportion of patients alive and progression free at 12 and 18 months	To match TFL standards
Section 4.2.2.3 Duration of Response	Removed analysis of Duration of Response	Not required
Section 4.2.3.3	Added mixed models repeated measures for analyses of PRO endpoints	To explore the treatment effect using change from baseline data

Section 3.2.1.2, 3.2.2.2, 3.2.2.7	Clarify criteria for censoring in the setting of no progression or death from last evaluable RECIST 1.1 assessment to last RECIST 1.1 assessment	*See below

*Rationale for the change:

Patient entry into the study is restricted to locally advanced, stage III NSCLC who have received definitive concurrent chemoradiation. These indicates that all patients would have localized diseases that have been radiated. It is known that localized post-radiation changes may affect lesion sizes and there may be high levels of necrosis/fibrosis with little or no active tumor in recently irradiated lesions. However, accepting these limitations in this patient population with prior curative radiation treatment, the prior irradiated lesions may be considered measurable at baseline and selected as target lesions provided they fulfil the other criteria for measurability according to RECIST 1.1. It is anticipated that a proportion of patients entering the study may have RECIST assessments being non-evaluable (NE) on-treatment, even if they have been followed up for multiple assessments and have not shown any evidence of PD, based on either their target, non-target or presence of any new lesion.

Therefore, in order to utilize all available tumor assessment follow-up information in the primary analysis of PFS, patients without a radiographic disease progression will be censored at their last assessment date, instead of the last evaluable assessment date.

7. REFERENCES

Berry et al 1991
Berry G, Kitchin RM, Mock PA. A comparison of two simple hazard ratio estimators based on the logrank test. Stat Med 1991 May;10(5):749-55.

Breslow, 1974
Breslow, NE. Covariance Analysis of Censored Survival Data. Biometrics 1974; 30:89–99.

Burman et al 2009
Burman CF, Sonesson C, Guilbaud O. A recycling framework for the construction of Bonferroni-based multiple tests. Stat Med 2009 Feb 28;28(5):739-61.

Butts et al 2014
Butts C, Socinski MA, Mitchell PL, Thatcher N, Havel L, Krzakowski M, et al, on behalf of the START trial team. Tecemotide (L-BLP25) vs placebo after chemoradiotherapy for stage III non-small-cell lung cancer (START): a randomised, double-blind, phase 3 trial. Lancet Oncol 2014;15(1):59-68.

Cockcroft and Gault 1976
Cockcroft DW, Gault MH. Prediction of creatinine clearance from serum creatinine. Nephron 1976; 16(1):31-41.

Collett 2003
Collett D. Modelling survival data in medical research: 2nd ed. Chapman and Hall/CRC. 2003. Available on request.

Ellis et al 2008
Ellis S, Carroll KJ, Pemberton K. Analysis of duration of response in oncology trials. Contemp Clin Trials 2008 Jul;29(4):456-65.

Fayers et al 1999
Fayers P, Aaronson NK, Bjordal K, Curran D, Groenvold M on behalf of the EORTC Quality of Life Study Group. EORTC QLQ-C30 Scoring Manual: 2nd Edition 1999. Available on request.

Fleischer et al 2011
Fleischer F, Gaschler-Markefski B, Bluhmki E. How is retrospective independent review influenced by investigator-introduced informative censoring: a quantitative approach. Stat Med 2011 Dec 20;30(29):3373-86.

Gail and Simon 1985
Gail M, Simon R. Tests for qualitative interactions between treatment effects and patient subsets. Biometrics 1985 Jun;41(2):361–72.

Glimm et al 2009
Glimm E, Maurer W, Bretz F. Hierarchical testing of multiple endpoints in group-sequential trials. Statistics in Medicine 2009; 29:219–28.

Hanna et al 2004
Hanna N, Shepherd FA, Fossella FV, Pereira JR, De Marinis F, von Pawel J, et al. Randomized phase III trial of pemetrexed vs docetaxel in patients with non-small-cell lung cancer previously treated with chemotherapy. J Clin Oncol 2004 May 1;22(9):1589-97.

Klein et al 2007
Klein JP, Logan B, Harhoff M, Andersen PK. Analyzing survival curves at a fixed point in time. Stat Med 2007 Oct 30; 26(24): 4505–19.

Lan and DeMets 1983
Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika 1983;70:659-63.

Nishino et al 2013
Nishino M, Giobbie-Hurder A, Gargano M, Suda M, Ramaiya NH and Hodi FS. Developing a common language for tumour response to immunotherapy: Immune-related response criteria using unidimensional measurements. Clin Cancer Res 2013;19:3936-43

Oemar and Janseen 2013
Oemar M, Oppe M. EQ-5D-3L User Guide: Basic information on how to use the EQ-5D-3L instrument V5.0 (October 2013). Available from URL: http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/EQ-5D-3L_UserGuide_2013_v5.0_October_2013.pdf. Accessed 07 January 2014.

Osoba et al 1998
Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol 1998 Jan;16(1):139-44.

Pintilie M.
Competing risks: A practical perspective. Wiley.

Robins and Tsiatis 1991
Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Communications in Statistics-Theory and Methods 1991; 20(8):2609-31.

Robins 1993
Robins JM. Information Recovery and Bias Adjustment in Proportional Hazards Regression Analysis of Randomized Trials Using Surrogate Markers. Proceedings of the Biopharmaceutical Section, American Statistical Association 1993; 24-33.

Sellke and Siegmund 1983
Sellke T, Siegmund D. Sequential analysis of the proportional hazards model. Biometrika 1983;70:315-26.

Stone et al 2015
Stone A, Macpherson E, Smith A, Jennison C. Model free audit methodology for bias evaluation of tumour progression in oncology. Submitted.

Sun and Chen 2010
Sun X, Chen C. Comparison of Finkelstein's Method with the Conventional Approach for Interval-Censored Data Analysis. Stat Biopharm Res 2010;2(1):97-108.

Whitehead and Whitehead 1991
Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomized clinical trials. Statistics In Medicine 1991;10:1665-77.

Wolchok et al 2009
Wolchok JD, Hoos A, O'Day S, Weber JS, Hamid O, Lebbé C, et al. Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. Clin Cancer Res 2009 Dec 1;15(23):7412-20.

Ye et al 2012

Ye Y, Li A, Lui L, Yao B. A group sequential Holm procedure with multiple primary endpoints. Statistics in Medicine 2012; 32:1112-1124

AstraZeneca

8. APPENDIX – DETAILS FOR TTDM

The locations of sites of distant metastases are shown in the table below (taken from Appendix 4.7 of the Export Requirements and Technical Specifications per irRECIST 1.1). Locations shown in grey will not be considered distant metastases.

Non-Nodal/New Non-Nodal Target Lesion Location (LESLOC)	Nodal Target/New Nodal Target Lesion Location (LESLOC)	Non-Target/New Lesion Lesion Location (NLESLOC)	New Non-Measurable Disease (NMESLOC)	AZ Location	AZ Code
			Liver		160
		Liver - multiple segments - right and left lobes		Liver - Multiple Segments - Right and Left Lobes	1
		Liver - multiple segments - left lobe		Liver - Multiple Segments - Left Lobe	2
		Liver - multiple segments - right lobe		Liver - Multiple Segments - Right Lobe	3
Liver - caudate lobe, segment I		Liver - caudate lobe, segment I		Liver - Caudate Lobe, Segment I	4
Liver - left lobe, segment II		Liver - left lobe, segment II		Liver - Left Lobe, Segment II	5
Liver - left lobe, segment III		Liver - left lobe, segment III		Liver - Left Lobe, Segment III	6
Liver - left lobe, segment IV		Liver - left lobe, segment IV		Liver - Left Lobe, Segment IV	7
Liver - right lobe, segment V		Liver - right lobe, segment V		Liver - Right Lobe, Segment V	8
Liver - right lobe, segment VI		Liver - right lobe, segment VI		Liver - Right Lobe, Segment VI	9
Liver - right lobe, segment VII		Liver - right lobe, segment VII		Liver - Right Lobe, Segment VII	10
Liver - right lobe,		Liver - right lobe, segment		Liver - Right Lobe, Segment VIII	11

segment VIII		VIII			
Liver - left lobe		Liver - left lobe		Liver - Left Lobe	12
Liver - right lobe		Liver - right lobe		Liver - Right Lobe	13
Liver - specify		Liver – specify		Liver, Specify	125
			Lung		159
		Lung - multiple lobes bilaterally		Lung - Multiple Lobes Bilaterally	14
		Lung - multiple lobes left		Lung - Multiple Lobes Left	15
		Lung - multiple lobes right		Lung - Multiple Lobes Right	16
Lung - left upper lobe		Lung - left upper lobe		Lung - Left Upper Lobe	17
Lung - left lower lobe		Lung - left lower lobe		Lung - Left Lower Lobe	18
Lung - right upper lobe		Lung - right upper lobe		Lung - Right Upper Lobe	19
Lung - right medial lobe		Lung - right medial lobe		Lung - Right Medial Lobe	20
Lung - right lower lobe		Lung - right lower lobe		Lung - Right Lower Lobe	21
Lung - left		Lung – left		Lung - Left	22
Lung - right		Lung – right		Lung - Right	23
Lung - specify		Lung – specify		Lung, Specify	126
			Lymph nodes		158
		Lymph nodes - bilateral cervical		Lymph Nodes - Bilateral Cervical	24
		Lymph nodes - bilateral hilar and mediastinal		Lymph Nodes - Bilateral Hilar and Mediastinal	25
		Lymph nodes - bilateral supraclavicular		Lymph Nodes - Bilateral Supraclavicular	26
		Lymph nodes - bilateral axillary		Lymph Nodes - Bilateral Axillary	27
		Lymph nodes - bilateral aorto-iliac		Lymph Nodes - Bilateral Aorto-Iliac	28
		Lymph nodes - bilateral pelvic		Lymph Nodes - Bilateral Pelvic	29

		Lymph nodes - bilateral inguinal		Lymph Nodes - Bilateral Inguinal	30
	cervical left	Lymph nodes - cervical left		Lymph Nodes - Cervical Left	31
	cervical right	Lymph nodes - cervical right		Lymph Nodes - Cervical Right	32
	supraclavicular left	Lymph nodes - supraclavicular left		Lymph Nodes - Supraclavicular Left	33
	supraclavicular right	Lymph nodes - supraclavicular right		Lymph Nodes - Supraclavicular Right	34
	axillary left	Lymph nodes - axillary left		Lymph Nodes - Axillary - Left	35
	axillary right	Lymph nodes - axillary right		Lymph Nodes - Axillary - Right	36
	mediastinal	Lymph nodes - mediastinal		Lymph Nodes - Mediastinal	37
	para-tracheal left	Lymph nodes - para-tracheal left		Lymph Nodes - Para-Tracheal - Left	38
	para-tracheal right	Lymph nodes - para-tracheal right		Lymph Nodes - Para-Tracheal - Right	39
	aortopulmonary window	Lymph nodes - aortopulmonary window		Lymph Nodes - Aortopulmonary Window	40
	hilar left	Lymph nodes - hilar left		Lymph Nodes - Hilar Left	41
	hilar right	Lymph nodes - hilar right		Lymph Nodes - Hilar Right	42
	subcarinal	Lymph nodes - subcarinal		Lymph Nodes - Subcarinal	43
	peri-cardiac	Lymph nodes - peri-cardiac		Lymph Nodes - Peri-Cardiac	44

	retrocrural	Lymph nodes - retrocrural		Lymph Nodes - Retrocrural	45
	porta hepatis	Lymph nodes - porta hepatis		Lymph Nodes - Porta Hepatis	46
	celiac axis	Lymph nodes - celiac axis		Lymph Nodes - Celiac Axis	47
	portocaval	Lymph nodes - portocaval		Lymph Nodes - Portocaval	48
	peripancreatic	Lymph nodes - peripancreatic		Lymph Nodes - Peripancreatic	49
	mesenteric	Lymph nodes - mesenteric		Lymph Nodes - Mesenteric	50
	peri-aortic	Lymph nodes - peri-aortic		Lymph Nodes - Peri-Aortic	51
	peri-caval	Lymph nodes - peri-caval		Lymph Nodes - Peri-Caval	52
	retroperitoneal	Lymph nodes - retroperitoneal		Lymph Nodes - Retroperitoneal	53
	iliac left	Lymph nodes - iliac left		Lymph Nodes - Iliac Left	54
	iliac right	Lymph nodes - iliac right		Lymph Nodes - Iliac Right	55
	pelvic left	Lymph nodes - pelvic left		Lymph Nodes - Pelvic Left	56
	pelvic right	Lymph nodes - pelvic right		Lymph Nodes - Pelvic Right	57
	inguinal left	Lymph nodes - inguinal left		Lymph Nodes - Inguinal Left	58

	inguinal right	Lymph nodes - inguinal right		Lymph Nodes - Inguinal Right	59
	Lymph nodes - specify	Lymph nodes - specify		Lymph Nodes, Specify	127
			Bone		161
Bone - specify		Bone – specify		Bone, Specify	128
		Bones - specify		Bones, Specify	129
		Bones - multiple axial skeleton		Bones - Multiple Axial Skeleton	60
		Bones - multiple appendicular skeleton		Bones - Multiple Appendicular Skeleton	61
		Pleural fluid - bilateral		Pleural Fluid - Bilateral	62
		Pleural fluid - left		Pleural Fluid - Left	63
		Pleural fluid - right		Pleural Fluid - Right	64
Pleura - bilateral		Pleura - bilateral		Pleura - Bilateral	65
Abdominal wall - left		Abdominal wall - left		Abdominal Wall - Left	66
Abdominal wall – anterior		Abdominal wall - anterior		Abdominal Wall - Anterior	67
Abdominal wall - right		Abdominal wall - right		Abdominal Wall - Right	68
Abdominal wall - posterior		Abdominal wall - posterior		Abdominal Wall - Posterior	69
Adrenal gland - left		Adrenal gland - left		Adrenal Gland - Left	70
Adrenal gland - right		Adrenal gland - right		Adrenal Gland - Right	71
Biliary tract		Biliary tract		Biliary Tract	72
Bladder		Bladder		Bladder	73
Bone- Skull/facial				Bone - Skull/Facial	144
Bone-Ribs, right				Bone - Ribs, Right	145
Bone- Ribs, left				Bone - Ribs, Left	146
Bone- Sternum				Bone - Sternum	147
Bone- Clavicle, right				Bone - Clavicle, Right	148

Bone- Clavicle, left				Bone - Clavicle, Left	149
Bone- Scapula, right				Bone - Scapula, Right	150
Bone- Scapula, left				Bone - Scapula, Left	151
Bone- Spine/Sacrum/Coccyx				Bone - Spine/Sacrum/Coccyx	152
Bone- Pelvis				Bone - Pelvis	153
Bone- upper extremity, right				Bone - Upper Extremity, Right	154
Bone- upper extremity, left				Bone - Upper Extremity, Left	155
Bone- lower extremity, right				Bone - Lower Extremity, Right	156
Bone- lower extremity, left				Bone - Lower Extremity, Left	157
			Brain		162
		Brain - bilateral		Brain - Bilateral	74
		Brain - brain stem		Brain - Brain Stem	75
		Brain - left cerebellum		Brain - Left Cerebellum	76
		Brain - left cerebral hemisphere		Brain - Left Cerebral Hemisphere	77
		Brain - right cerebellum		Brain - Right Cerebellum	78
		Brain - right cerebral hemisphere		Brain - Right Cerebral Hemisphere	79
		Brain – specify		Brain, Specify	130
Breast - left		Breast – left		Breast - Left	80
Breast - right		Breast – right		Breast - Right	81
Chest wall - anterior		Chest wall - anterior		Chest Wall - Anterior	84
Chest wall - left		Chest wall - left		Chest Wall - Left	85
Chest wall - posterior		Chest wall - posterior		Chest Wall - Posterior	86
Chest wall - right		Chest wall - right		Chest Wall - Right	87

Colon		Colon		Colon	88
		Diaphragm - bilateral		Diaphragm - Bilateral	89
Diaphragm - left		Diaphragm - left		Diaphragm - Left	90
Diaphragm - right		Diaphragm - right		Diaphragm - Right	91
Heart		Heart		Heart	92
Kidney - left		Kidney – left		Kidney - Left	93
Kidney - right		Kidney – right		Kidney - Right	94
Mediastinum		Mediastinum		Mediastinum	95
Neck		Neck		Neck	96
Oesophagus		Oesophagus		Oesophagus	97
Ovary - left		Ovary – left		Ovary - Left	98
Ovary - right		Ovary – right		Ovary - Right	100
Pancreas		Pancreas		Pancreas	101
		Pelvis		Pelvis	102
Pelvis - left		Pelvis - left		Pelvis - Left	103
Pelvis - central		Pelvis - central		Pelvis - Central	104
Pelvis - right		Pelvis – right		Pelvis - Right	105
Pericardium		Pericardium		Pericardium	106
		Peritoneal fluid (ascites) – left		Peritoneal Fluid (Ascites) - Left	107
		Peritoneal fluid (ascites) - right		Peritoneal Fluid (Ascites) - Right	108
		Peritoneal fluid (ascites) - cul-de-sac		Peritoneal Fluid (Ascites) - Cul-de-sac	109
		Peritoneal fluid (ascites) - all quadrants		Peritoneal Fluid (Ascites) - All Quadrants	110
Peritoneum/Omentum - central		Peritoneum/Omentum - central		Peritoneum/Omentum - Central	111
Peritoneum/Omentum - left		Peritoneum/Omentum – left		Peritoneum/Omentum - Left	112

Peritoneum/Omentum - right		Peritoneum/Omentum – right		Peritoneum/Omentum - Right	113
		Peritoneum/Omentum - all quadrants		Peritoneum/Omentum - All Quadrants	114
Pleura - left		Pleura – left		Pleura - Left	115
Pleura - right		Pleura – right		Pleura - Right	116
Prostate		Prostate		Prostate	117
Retroperitoneum		Retroperitoneum		Retroperitoneum	118
Skin - specify		Skin – specify		Skin, Specify	119
			Soft tissue/Skin		163
Small intestine		Small intestine		Small Intestine	120
Spleen		Spleen		Spleen	121
Stomach		Stomach		Stomach	122
Tracheal – Bronchial Tree		Tracheal – Bronchial Tree		Tracheal - Bronchial Tree	123
Uterus		Uterus		Uterus	124
Other, specify		Other, specify		Other, Specify	99
			Other		164